

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 7, 2002, 04:26:35 ; Search time 1869.5 Seconds
(without alignments)
3487.044 Million cell updates/sec

Title: US-09-904-568-2
Perfect score: 1112
Sequence: 1 MSALGAAPYLLHPADSHG.....ALVLKEMAELEKCKERKLD 224

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp
-MODEL=frame+p2n.model -DEV-xlp
-Q/cgn2.1/USPTO.spool/US09904568/runat_25112002_143304_19923/app_query.fasta_1.782
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09904568 @CGN_1_1_3637_@runat_25112002_143304_19923 -NCPU=6 -ICPU=3
-NO XLPXY -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sv.*
39: em_hgtgo_hum.*
40: em_hgtgo_mus.*
41: em_hgtgo_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1112	100.0	1100	10	AF290194 Rattus no
2	1103	99.2	97371	2	AC124883 Rattus no
3	1067	96.0	139408	2	AC079565 Mus muscu
4	1062	95.5	981	10	BC025891 Mus muscu
5	913	82.1	993	6	AX405845 Sequence
6	913	82.1	1030	9	BC002672 Homo sapi.
7	913	82.1	1124	9	BC003055 Homo sapi.
8	913	82.1	1355	9	AF290195 Homo sapi
9	913	82.1	3112	9	AK023070 Homo sapi
10	913	82.1	69587	9	AF146367 Homo sapi
11	913	82.1	127805	9	AC130271 Homo sapi
12	913	82.1	156503	9	AC084347 Homo sapi
13	913	82.1	183457	9	AF216665 Homo sapi
14	913	82.1	181858	2	AC084856 Homo sapi
15	913	82.1	186024	2	AC084225 Homo sapi
16	913	82.1	218445	2	AC083959 Homo sapi
17	913	82.1	345524	9	AF235103 Homo sapi
18	912	82.0	995	9	AK001477 Homo sapi
19	881	79.2	159779	2	AC105449 Homo sapi
20	862.5	77.6	139408	2	AC079565 Mus muscu
21	810.5	72.9	1588	9	AF113540 Homo sapi
22	792	71.2	90171	9	AC095040 Homo sapi
23	792	71.2	211844	2	AC048370 Homo sapi
24	680.5	61.2	70972	2	AC130332 Homo sapi
25	281	25.3	406	6	AX408617 Sequence
26	130	11.7	1569	3	AY122267 Drosophil
27	128.5	11.6	1566	10	BC034879 Mus muscu
28	126.5	11.4	39369	2	AC115681 Dictyoste
29	124.5	11.2	1093	9	AK002147 Homo sapi
30	124.5	11.2	1456	9	BC005179 Homo sapi
31	124.5	11.2	1462	6	AX048099 Sequence
32	113	10.2	18876	1	AE004835 Pseudomon
33	107.5	9.7	48000	2	AC020147 Drosophil
34	107.5	9.7	131266	10	AL669868 Mouse DNA
35	107.5	9.7	161754	3	AC007451 Drosophil
36	107.5	9.7	270055	3	AE003628 Drosophil
37	105.5	9.5	965	6	AX061981 Sequence
38	105	9.4	904	9	AF059618 Homo sapi
39	104	9.4	1098	6	AX061978 Sequence
40	104	9.4	20400	1	AE007068 Mycobacte
41	104	9.4	37586	1	MTCY270 Mycobacte
42	104	9.4	37586	6	AX191745 Sequence
43	104	9.4	149775	2	AC124338 Mus muscu
44	103.5	9.3	1362	6	AR169155 Sequence
45	103.5	9.3	1362	6	AR182445 Sequence

ALIGNMENTS

RESULT 1

AF290194 1100 bp mRNA linear ROD 10-OCT-2000
 LOCUS AF290194 Rattus norvegicus hypertension-related calcium-regulated gene mRNA,
 DEFINITION complete cds.
 ACCESSION AF290194
 VERSION AF290194.1 GI:9964284
 KEYWORDS
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus;
 Rattus;
 1 (bases 1 to 1100)
 Solban,N., Jia,H.-P., Richard,S., Tremblay,S., Devlin,A.M.,
 Peng,J., Gosard,F., Guo,D.-F., Movel,G., Hamet,P., Lewanczuk,R.
 and Tremblay,J.
 Hcarg, a novel calcium-regulated gene coding for a nuclear protein,
 is potentially involved in the regulation of cell proliferation
 J. Biol. Chem. 275 (41), 32234-32243 (2000)
 REFERENCE
 AUTHORS
 JOURNAL
 MEDLINE
 PUBMED
 20493613
 10918053
 2 (bases 1 to 1100)
 Tremblay,J., Solban,N., Lewanczuk,R., Jia,H.-P. and Hamet,P.
 Direct Submission
 Submitted (26-JUL-2000) CHUM Research Center, University of
 Montreal, St-Urbain, Montreal, Qc H2W 1H8, Canada
 FEATURES
 source
 1..1100
 location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="SHR"
 /db_xref="taxon:10116"
 /tissue_type="parathyroid"
 132..806
 /note="hypertension-related calcium-regulated gene Carg"
 /codon_start=1
 /product="unknown"
 /protein_id="AAG0914.1"
 /db_xref="GI:9964285"
 /translation="MSALGAAAPYLHPHDSGRVSLGSGPSEVTAVALKDD
 RSTPRKLIVGALHGKDEAVEQGLASNLSEBRALVLAGHTLLLOALPLPA
 SLKPAPEELQELIGIPDILIGDLASLAFSGORPLDSVAQGGSSLPVSYFPMRD
 VAISTASRSLSPLMQLKLTGSSAHRFEVPLAKFQELRYVALVKEMAELEKCC
 ERKLOD"
 BASE COUNT 260 a 279 c 315 g 246 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 9.35e-89 Length: 1100
 Score: 1112.00 Matches: 224
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-904-568-2 (1-224) x AF290194 (1-1100)
 QY 1 MetSerAlaLeuGAlaAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
 Db 132 ATGTCTGCTTGGGGGGCGAGCTCCATCTTGCACCTCCCGCTGACAGTCACAGTGGC 191
 QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40
 Db 192 CGGCTCACTTTCCTGGTCCAGCCCTCTCCAGAACTGACGGCGTGGCTCAGCTCTTG 251
 QY 41 LysAspLeuAspArgSerThrPheArgPheLeuLeuValValGlyAlaLeuHis 60
 Db 252 AAGACCTTAGCAGGAGACCTTCAGAAAGTTGTGAACCTTGAGTGGGGCCCTTCAT 311
 QY 61 GlyLysAspCysArgGluAlaValGlnGlnLeuGlyAlaSerAlaAsnLeuSerGln 80
 Db 312 GGGAAAGACTCCAGAGAGCTGTGAGCACTGTGTCACCCGCAACCTCTCAGAAAG 371
 QY 81 ArgLeuAlaValLeuLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100

Db 372 CGCTGCGCCGCTCTGTGCGGGACACACACCTCTCTCAGACGCTCTCCGGCTGCC 431
 QY 101 ProAlaSerLeuLeuProAlaPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
 Db 432 CCGTCAATCTTAAGCAACCAATGCTCTTCCAGGAAGAGCTCCAGGAACCTTGCTCAG 491
 QY 121 AspLeuLeuGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAsp 140
 Db 492 GATCTAATTTGAGATTTTGGCCAGTTTGGCATTTGGAGTCAACGCCCTTCTTCACTCT 551
 QY 141 ValAlaGlnGlnGlnGlySerSerLeuProHisValSerTyrPheArgTyrValAsp 160
 Db 552 GTAGCCCAACAGCAGAGATCTCGCTGCTCAGCGTCTTACTTCCGTTGGGGGTGAC 611
 QY 161 ValAlaIleSerThrSerAlaGlnSerTyrSerLeuGlnProSerValLeuMetGlnLeu 180
 Db 612 GTGGCACTTCAACAGCGCTCAGTCCGCTCCCGCAACCGAGTCTCATGACAGCTG 671
 QY 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaValPheGlnGlnLeu 200
 Db 672 AAGCTCACAGATGATCTGCACACCGCTTCCAGGTGGCCCAATTCACAGAGCTG 731
 QY 201 ArgTyrSerValAlaLeuValLeuGlyGlnMetAlaGlnLeuGlnGlnGlnGlnGln 220
 Db 732 CGGTACAGTGTAGCTTGTGCTTAAAGAGATGACAGAACTGAGAAAGAACTGTGAGCGC 791
 QY 221 LysLeuGlnAsp 224
 Db 792 AAACGTCAGGAC 803
 RESULT 2
 AC124883/c 97371 bp DNA linear HTG 31-JUL-2002
 LOCUS Rattus norvegicus clone CH230-335N13, *** SEQUENCING IN PROGRESS
 DEFINITION *** 48 unordered pieces.
 ACCESSION AC124883.1 GI:21465224
 VERSION HTG; HTGS PHASE1.
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 97371)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooke,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbarella,J., Benton,J., Bimge,K., Blenkinsburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Bhaya,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Duthin,K.D.,
 Earmhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homst,F., Howard,S., Huber,J., Huiyk,S., Hume,D., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousegod,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunnu,G.,

Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojokan, I., Roife, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, K., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 2 (bases 1 to 97371)
 Worley, K.C.
 Direct Submission
 Submitted (19-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 97371)
 Worley, K.C.
 Direct Submission
 Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KAXX
 Center clone name: CH230-335N13
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 59101 bases at least Q40
 Consensus quality: 63071 bases at least Q30
 Consensus quality: 66970 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 42 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1183: contig of 1183 bp in length
 * 1184 1283: gap of unknown length
 * 1284 2784: contig of 1501 bp in length
 * 2785 2884: gap of unknown length
 * 2885 4397: contig of 1513 bp in length
 * 4398 4497: gap of unknown length
 * 4498 5544: contig of 1047 bp in length
 * 5545 7134: gap of unknown length
 * 7135 7234: contig of 1490 bp in length
 * 7235 8472: gap of unknown length
 * 8473 8573: contig of 1238 bp in length
 * 8574 9817: gap of unknown length
 * 9818 11055: contig of 1245 bp in length
 * 11056 12327: contig of 1138 bp in length
 * 12328 12427: gap of unknown length
 * 12428 13858: contig of 1172 bp in length
 * 13859 13959: gap of unknown length
 * 13959 15185: contig of 1431 bp in length
 * 15185 15185: contig of 1227 bp in length

15186 15285: gap of unknown length
 15286 16945: contig of 1660 bp in length
 16946 17045: gap of unknown length
 17046 19195: contig of 2150 bp in length
 19196 19295: gap of unknown length
 20580 20580: contig of 1285 bp in length
 20581 22355: gap of unknown length
 22356 22455: contig of 1675 bp in length
 22456 24323: gap of unknown length
 24324 24432: contig of 1868 bp in length
 24433 25682: gap of unknown length
 25683 25782: contig of 1259 bp in length
 25783 28074: gap of unknown length
 28075 28174: contig of 2292 bp in length
 28175 29504: gap of unknown length
 29505 29604: contig of 1330 bp in length
 29605 32241: gap of unknown length
 32242 32341: contig of 2637 bp in length
 32342 33911: gap of unknown length
 33912 34011: contig of 1570 bp in length
 34012 35064: gap of unknown length
 35065 35164: contig of 1053 bp in length
 35165 37991: gap of unknown length
 37992 40938: contig of 2727 bp in length
 40939 41038: gap of unknown length
 41039 43235: contig of 2947 bp in length
 43236 43335: gap of unknown length
 43336 47102: contig of 2197 bp in length
 47103 50377: gap of unknown length
 50378 50476: contig of 3667 bp in length
 50477 54251: gap of unknown length
 54252 54351: contig of 3274 bp in length
 54352 56883: gap of unknown length
 56884 56983: contig of 3775 bp in length
 56984 59558: gap of unknown length
 59559 62434: contig of 2532 bp in length
 62435 62535: gap of unknown length
 62536 64311: contig of 2776 bp in length
 64312 64411: gap of unknown length
 64412 66566: contig of 1777 bp in length
 66567 66666: gap of unknown length
 66667 68269: contig of 2155 bp in length
 68270 71250: gap of unknown length
 71251 71351: contig of 1803 bp in length
 71352 74840: gap of unknown length
 74841 77886: contig of 4844 bp in length
 77887 77987: gap of unknown length
 77988 84511: contig of 2881 bp in length
 84512 85551: gap of unknown length
 85552 85667: contig of 3490 bp in length
 85668 85767: gap of unknown length
 85768 90611: contig of 2946 bp in length
 90612 90711: gap of unknown length
 90712 94028: contig of 4316 bp in length
 94029 94127: gap of unknown length
 94128 97371: contig of 3244 bp in length.
 ----- Location/Qualifiers
 1. 97371
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-335N13"
 BASE COUNT 23717 a 20869 c 20881 g 23584 t 8320 others
 ORIGIN
 Alignment Scores: 1.64e-85 Length: 97371
 Pred. No.: 1103.00 Matches: 221
 Score:

Percent Similarity: 99.55%
 Best Local Similarity: 98.66%
 Query Match: 99.19%
 DB: 2
 Conservative: 2
 Mismatches: 1
 Indels: 0
 Gaps: 0

US-09-904-568-2 (1-224) x AC124883 (1-97377)

QY 1 MetSerAlaLeuG1yAlaAlaAlaProTYrLeuH1sProAlaAspSerH1sSerg1y 20
 DB 62064 ATCTGCTTTGGGCGCTCAGCTCATCTTCCACCATCCGCTGACAGTCACTGGC 62005
 QY 21 ArgValSerPheLeuG1ySerG1nProSerProG1uValThrAlaValaG1nLeu 40
 DB 62004 CGGCTAGTTTCTGGGCTTCCAGCCCTTCCAGAACTGACGCGCTGCTGCTTGG 61945
 QY 41 LysAspLeuAspArgSerThrPheArgLysLeuLysLeuValaG1yAlaLeuH1s 60
 DB 61944 AAGGACTTGCAGAGCACTTCCAGAACTTGTAACTTGTAGTCCGGGCTTGCAT 61885
 QY 61 GlyLysAspCysArgG1uAlaValaG1nLeuG1yAlaSerAlaAsnLeuSerg1u 80
 DB 61884 CGGAAAGACTGCAGAGAACTGTGAGCACTTGTGCTGCGCCCACTGTCAAGAG 61825
 QY 81 ArgLeuAlaValaLeuLeuAlaG1yThrH1sThrLeuLeuG1nAlaLeuArgLeuPro 100
 DB 61824 CGCTGGCCCTCTGCTGCGGCGACACACCTGCTTCCAGAGCTTCCGGCTGCC 61765
 QY 101 ProAlaSerLeuLysProAspAlaPheG1nG1uLeuG1nG1uLeuG1yLeuProG1n 120
 DB 61764 CCGTCTAGCTTAAAGCCACATCTCTTCCAGAAAGCTTCCAGAACTTGGCTTCTCAG 61705
 QY 121 AspleuL1egL1yAspLeuAlaSerLeuAlaPheG1ySerg1nArgProLeuLeuAspSer 140
 DB 61704 GATCTAATTGGAGATTGGCCGATTGGCATTTGGAGTCAACGCGCTCTTCTGACATCT 61645
 QY 141 ValAlaG1nG1nG1ySerSerLeuProH1sValSerTYrPheArgPheArgValaAsp 160
 DB 61644 CTAGCCCAACAGCAGGATCTCTGCTGCTTCCATGCTTCTTCCGGGGGGGGGAGC 61585
 QY 161 ValAlaL1eSerThSerAlaG1nSerArgSerLeuG1nProSerValLeuLeuG1nLeu 180
 DB 61584 GTGGCCATCTCAACAGCGCTCAGTCCGCTCTTCCAGAAAGCTTCTTCAAGAGCTG 61525
 QY 181 LysLeuThrAspG1ySerAlaH1sArgPheG1uValProL1eAlaLysPheG1nG1uLeu 200
 DB 61524 AAGCTCAAGATGATCTGCACACGCTTCCAGGTCGCTTCCAAATTCAGAGAGCTG 61465
 QY 201 ArgTYrSerValaLeuValaLeuValaLeuValaLeuValaLeuValaLeuValaLeu 220
 DB 61464 CGGTACACTGTAGCTTATCTTATAGAGATGCGAAGCTGAGAAAGATGTAGAGCGC 61405
 QY 221 LysLeuG1nAsp 224
 DB 61404 AAACGACAGGAC 61393

RESULT 3
 LOCUS AC079565/c 139408 bp DNA linear HTG 02-SEP-2000
 DEFINITION Mus musculus clone RP23-55E9, WORKING DRAFT SEQUENCE, 27 unordered
 pieces.
 AC079565
 VERSION AC079565.1 GI:9964930
 KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 139408)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Mouse
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 139408)
 AUTHORS DOE Joint Genome Institute.

TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

Project Information
 Center Project Name: 1758097
 Center Clone Name: RPCI-23_55E9

Summary Statistics

Consensus quality: 117884 bases at least Q40
 Consensus quality: 128049 bases at least Q30
 Consensus quality: 130435 bases at least Q20
 Estimated insert size: 221000; agarose-fp estimation
 Estimated insert size: 136808; sum-of-contigs estimation
 Quality coverage: 9.0 in Q20 bases; agarose-fp estimation
 Quality coverage: 14.53 in Q20 bases; sum-of-contigs estimation
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1412: contig of 1412 bp in length
 1413 1512: gap of unknown length
 1513 2645: contig of 1133 bp in length
 2646 2745: gap of unknown length
 2746 3772: contig of 1027 bp in length
 3773 3872: gap of unknown length
 3873 5043: contig of 1171 bp in length
 5044 5144: gap of unknown length
 5144 6289: contig of 1146 bp in length
 6290 6390: gap of unknown length
 6390 8113: contig of 1724 bp in length
 8114 8213: gap of unknown length
 8214 10028: contig of 1815 bp in length
 10029 10128: gap of unknown length
 10129 12184: contig of 2056 bp in length
 12185 12284: gap of unknown length
 12285 14305: contig of 2021 bp in length
 14306 14405: gap of unknown length
 14406 16673: contig of 2268 bp in length
 16674 16773: gap of unknown length
 16774 17818: contig of 1045 bp in length
 17819 17918: gap of unknown length
 17919 20432: contig of 2514 bp in length
 20433 20532: gap of unknown length
 20533 22317: contig of 1785 bp in length
 22318 22417: gap of unknown length
 22418 24449: contig of 2032 bp in length
 24450 24549: gap of unknown length
 24550 29501: contig of 4952 bp in length
 29502 29601: gap of unknown length
 29602 34397: contig of 4796 bp in length
 34398 34497: gap of unknown length
 34498 40960: contig of 6463 bp in length
 40961 41060: gap of unknown length
 41061 48574: contig of 7514 bp in length
 48575 48674: gap of unknown length
 48675 56444: contig of 7770 bp in length
 56445 56544: gap of unknown length
 56545 63595: contig of 7051 bp in length
 63596 71636: gap of unknown length
 71637 71736: gap of unknown length
 71737 79381: contig of 7645 bp in length
 79382 79481: gap of unknown length
 79482 85904: contig of 6423 bp in length

* 85905 86004: gap of unknown length
* 86005 94184: contig of 8180 bp in length
* 94185 94284: gap of unknown length
* 94285 103274: contig of 8990 bp in length
* 103275 103374: gap of unknown length
* 103375 116417: contig of 13043 bp in length
* 116418 116517: gap of unknown length
* 116518 139408: contig of 22891 bp in length.

FEATURES
source
1. 139408
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-55E9"
/clone_lib="RFC1 mouse BAC library 23"
BASE COUNT 36138 a 32868 c 32375 g 35393 t 2634 others
ORIGIN

Alignment Scores:
Pred. No.: 3,76e-82 Length: 139408
Score: 1067.00 Matches: 215
Percent Similarity: 97.77% Conservative: 4
Best Local Similarity: 95.98% Mismatches: 5
Query Match: 95.95% Indels: 0
DB: 2 Gaps: 0

US-09-904-568-2 (1-224) x AC079565 (1-139408)

QY 1 MetSerAlaLeuGlyAlaAlaProTyrLeuHisHisProAlaAspSerHisSergly 20
Db 137638 ATGCTCGGTGGGGCTGGCGCTCCATATTACACATCCAACTGACAGTCACAGCGGC 137579
QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40
Db 137578 CGGGTCAGTTCTCGGGTTCCTCCAGCCCTTCGAGAGTGCAGCGAGTGGCCAGCTCTTG 137519
QY 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db 137518 AAGACTTAGACAGGACACCTTCAGAACTGTTGAACTCGTAGCGGGCCCTCCAT 137459
QY 61 GlyLysAspCysArgGluAlaValGluGlnLeuGlyAlaSerAlaSerLeuSerGluGlu 80
Db 137458 GGGAAACACTGCAGAGAGCGGCTGGTGGCCAGCACTTGGTCCAGCGCAACCTGTGAGAGGAG 137399
QY 81 ArgLeuAlaValLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 137398 CGTCTGCCGCTCTGCTGGCGGGCACACACCTGCTCCAGCAGGCTCTCCGGCTGCC 137339
QY 101 ProAlaSerLeuLysProAspAlaPheGlnGlnGluLeuGlnLeuGlyIleProGln 120
Db 137338 CCTGCCAGTCTGAACCCAGATGCTTCCAGGACGAACTCCAGGAACCTTGCATTCCCAA 137279
QY 121 AspLeuIleGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db 137278 GATATGATTGGAGATCTGGCCAGTTTGGCATTTGGGAGTCAAGCCCTCTTCTCGACTCT 137219
QY 141 ValAlaGlnGlnGlnGlySerSerLeuProHisValSerTyrPheArgTrpArgValAsp 160
Db 137218 GTAGCCACAGCAGAGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 137159
QY 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 137158 GTGGCCATCTCAACAGTGCTCAGTCCCGCTCCCTGCAACCGAGTGTCTCATGAGCTG 137099
QY 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
Db 137098 AAGCTTCACAGATGGATCGGCACACCGCTTCGAGGTGCGCATAGCAAAATCCAGAGCTG 137039
QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGlyLysCysGluArg 220
Db 137038 CGCTACGGCTGGCTTGGTCTTTAAGAGATGGCAGAGCTGGAGAGAAAGTGTGAGCGC 136979
QY 221 LysLeuGlnAsp 224

Db 136978 AAATGCGAGGAC 136967

RESULT 4
BC025891
LOCUS
DEFINITION
IMAGE:5035059, mRNA, complete cds.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

NTI-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 45 Row: h Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source
1. 981
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZECH II"
/clone="MGC:32364 IMAGE:5035059"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
71..745
/codon_start=1
/product="RIKEN CDNA 2310065H03 gene"
/protein_id="AAH25891.1"
/db_xref="GI:19484145"
/db_xref="LocusID:66398"
/translation="MSALGAPAPVYLHPTDSHSGRVSLGSPSAEVTAVALLKDL
RSTFRKLIVVGLHKGKREAVQHLGASNLSERLAVLACTHTLLQOALRLPPA
SLKPDADFDELQELGIPDMIGLAFSGQRELLSSVAQQQSSILPRVSNFWRVD
VAISTAQSRSLQPSVLMLQKLTDGSAHRFVPIAKFOELRYSVALVLKMAELERK
ERKLQD"

CDS

BASE COUNT 231 a 254 c 284 g 212 t
ORIGIN

Alignment Scores:
Pred. No.: 2e-84 Length: 981
Score: 1062.00 Matches: 214
Percent Similarity: 97.32% Conservative: 4
Best Local Similarity: 95.54% Mismatches: 6

Query Match:	95.50%	Indels:	0
DB:	10	Gaps:	0
US-09-904-568-2 (1-224) x BC025891 (1-981)			

```

/db_xref="taxon:9606"
98.772
CDS

```

BASE COUNT	205 a	280 c	299 g	209 t
ORIGIN				

Alignment scores:	
Pred. No.:	2,49e-71
Score:	913.00
Percent Similarity:	90.18%
Best local Similarity:	80.36%
Query Match:	82.10%
DB:	6
Length:	992
Matches:	180
Conservative:	22
Mismatches:	22
Indels:	0
Gaps:	0

US-09-904-568-2 (1-224) X AX405845 (1-993)

1 messeiatleuglalaalalaprotyrleuhishisproalaspserhissergly 20
 28
 36
 44
 52
 60
 68
 76
 84
 92
 100
 108
 116
 124
 132
 140
 148
 156
 164
 172
 180
 188
 196
 204
 212
 220
 228
 236
 244
 252
 260
 268
 276
 284
 292
 300
 308
 316
 324
 332
 340
 348
 356
 364
 372
 380
 388
 396
 404
 412
 420
 428
 436
 444
 452
 460
 468
 476
 484
 492
 500
 508
 516
 524
 532
 540
 548
 556
 564
 572
 580
 588
 596
 604
 612
 620
 628
 636
 644
 652
 660
 668
 676
 684
 692
 700
 708
 716
 724
 732
 740
 748
 756
 764
 772
 780
 788
 796
 804
 812
 820
 828
 836
 844
 852
 860
 868
 876
 884
 892
 900
 908
 916
 924
 932
 940
 948
 956
 964
 972
 980
 988
 996
 1004
 1012
 1020
 1028
 1036
 1044
 1052
 1060
 1068
 1076
 1084
 1092
 1100
 1108
 1116
 1124
 1132
 1140
 1148
 1156
 1164
 1172
 1180
 1188
 1196
 1204
 1212
 1220
 1228
 1236
 1244
 1252
 1260
 1268
 1276
 1284
 1292
 1300
 1308
 1316
 1324
 1332
 1340
 1348
 1356
 1364
 1372
 1380
 1388
 1396
 1404
 1412
 1420
 1428
 1436
 1444
 1452
 1460
 1468
 1476
 1484
 1492
 1500
 1508
 1516
 1524
 1532
 1540
 1548
 1556
 1564
 1572
 1580
 1588
 1596
 1604
 1612
 1620
 1628
 1636
 1644
 1652
 1660
 1668
 1676
 1684
 1692
 1700
 1708
 1716
 1724
 1732
 1740
 1748
 1756
 1764
 1772
 1780
 1788
 1796
 1804
 1812
 1820
 1828
 1836
 1844
 1852
 1860
 1868
 1876
 1884
 1892
 1900
 1908
 1916
 1924
 1932
 1940
 1948
 1956
 1964
 1972
 1980
 1988
 1996
 2004
 2012
 2020
 2028
 2036
 2044
 2052
 2060
 2068
 2076
 2084
 2092
 2100
 2108
 2116
 2124
 2132
 2140
 2148
 2156
 2164
 2172
 2180
 2188
 2196
 2204
 2212
 2220
 2228
 2236
 2244
 2252
 2260
 2268
 2276
 2284
 2292
 2300
 2308
 2316
 2324
 2332
 2340
 2348
 2356
 2364
 2372
 2380
 2388
 2396
 2404
 2412
 2420
 2428
 2436
 2444
 2452
 2460
 2468
 2476
 2484
 2492
 2500
 2508
 2516
 2524
 2532
 2540
 2548
 2556
 2564
 2572
 2580
 2588
 2596
 2604
 2612
 2620
 2628
 2636
 2644
 2652
 2660
 2668
 2676
 2684
 2692
 2700
 2708
 2716
 2724
 2732
 2740
 2748
 2756
 2764
 2772
 2780
 2788
 2796
 2804
 2812
 2820
 2828
 2836
 2844
 2852
 2860
 2868
 2876
 2884
 2892
 2900
 2908
 2916
 2924
 2932
 2940
 2948
 2956
 2964
 2972
 2980
 2988
 2996
 3004
 3012
 3020
 3028
 3036
 3044
 3052
 3060
 3068

.....CCGACATCCTGGTGAAGTCACAGTGGC 157

Db

158 CGAGTAGATTCTTGGGGGCCAGCTTCTCCAGAGGTGGCAGCAATGCCCCGGCTACTA 217

158 CGAGTGAATTCCTGGGGGCCAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 217

Db 218 GGGGACCTAGACAGGAGCACTTCAGAAAGTTCTCTGAAGTTGTGTCAGCAGCCTGCAG 2177

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Db 278 GGGGAGACTGCCGAGAGCGCTCTGGGGGTGAGCGCCACCTGCCGAGAG 337

218 GGGAGACIUGCAGAGGCTGTGGGGTCAGCGCCAACCTGCCGGAGGAG 337

338 CAGCTGGGGTGCCTGCTGCAGGCATGCACACACTGCTCCAGAGGCCCTCCGCTGCCC 397
 01 ArgLeuAlaValLeuLeuAlaGlyIleHisIleLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db

Db 338 CAGCTGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGTCGCCC 397

QY 101 ProLaserLeuLysProAspAlaIheGlnGluGluLeuGlnGlnGluLeuGlyIleProGln 120

Db 398 CCGACAGCCGGAAAGCCGACACCTTACGATCCCCAGCCTTCGCCACTTGCCTTTGTTTTTTT

[illegible]

OY 121 AspleuilegIAspleuAlaserleuAlapheglySerGlnargProleuleuaspSer 140
 ::::~::~:
 ::::~::~:
 ::::~::~:

[illegible]

QY 141 VALAAGINGINGLYSERSELEUPROHISVALSERTYRPHENARGTRPARCVALASP 160
::: :::

```

      ..::..
      yseiseluonhisvalsettyrheargtrpargValaspr 160

```

OY

ValAlaIleSerThrSerA laGlnSerArgSerLeu GInProSeVaLlMetGln Leu180	161
--	------------

.....TCCCTCAGATCGACATTGGTGCGGTGCAT577

161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180

DB 578 GTAGCAATCTCCACCAGTAGTCCCTCGCTCCCTCGACGCCGAGCTCCTGATGCAGCTG 637

QY 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaIleArgPheGlnGluLeu 300

Db	638	AAAGCTTTCAGATGGGCTCAGCATACCGCTTTAGGATCCCAAGTTCAGGAGCTG	697
AY	201	AAAGCTTTCAGATGGGCTCAGCATACCGCTTTAGGATCCCAAGTTCAGGAGCTG	

201 ArcticServer] All available - 201

Db 698 CGGTACAGCGTGGCCCTGGTCTCTAAAGGACATGGCAGATCTGGAGAAAGAGGTGTAGCCG 757

.....CCCTTTCATGACCAAGCGAAGCAGGGTTGTGCCTTGAGCTC /57

Db 758 AGACTGCAGGAC 769

RESULT 6

```
BC002672      1030 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS
DEFINITION      Homo sapiens, HT002 protein; hypertension-related calcium-regulated
                  gene, clone MGC:3418 IMAGE:3606279, mRNA, complete cds.
ACCESSION      BC002672
VERSION
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1030)
Strausberg, R.
Direct Submission
Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc\_mgc@hgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stancirip, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 12 Row: k Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7022758.
Location/Qualifiers
1. .1030
/organism="Homo sapiens"
/db_xref="LocusID:28991"
/db_xref="taxon:9606"
/clone="MGC:3418 IMAGE:3606279"
/tissue type="Uterus, endometrium adenocarcinoma"
/clone_lib="NIH MGC 44"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
126. .800
/codon_start=1
/product="HT002 protein; hypertension-related
calcium-regulated gene"
/protein_id="AAH02672.1"
/db_xref="GI:12803673"
/translations="MSA:GAATPYLHPGDSHGVSFLCAQLPEVAMARELLGLD
RSTFRLLKLVSSLOGEQREAVQRLGVSANLPBQLGALLAGHMTLQQALRLPPT
SLKPTDFRDLQELCIPODLVGLASVFGSQRPDLDSVAQQGAWLPHVADFWRVD
VAITSALARSQPSVLMQLKLDGSAYRFEVPTAKFQELRYSLVALVKEMADLEKRC
ERLQD"
BASE COUNT      218 a      288 c      313 g      211 t
ORIGIN
Alignment Scores:
Pred. No.:      2,61e-71      Length:      1030
Score:      913.00      Matches:      180
Percent Similarity:      90.18%      Conservative:      22
Best Local Similarity:      80.36%      Mismatches:      22
Query Match:      82.10%      Indels:      0
DB:      Gaps:      9
```

```
US-09-904-568-2 (1-224) x BC002672 (1-1030)
QY      1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
Db      126 ATGTCGTGCTGGGGGCTCAATCCATACCTGCATCATCTCTGGTGATAGTCACAGTGGC 185
QY      21 AtgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40
Db      186 CGAGTCAGTTCTTTGGGGGCCAGCTTCCTCCAGAGGTGCAGCAATGCCCGGTACTA 245
QY      41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db      246 GGGGAGCTACACAGGACGCTTCAGAAAGTTGCTGAAGTTGTGGTCAGCAGCCTGCAG 305
QY      61 GlyLysAspCysArgGluAlaValGlnLeuGlyAlaSerAlaAsnLeuSerGluGlu 80
Db      306 GGGGAGGAGCTGCCAGAGGCTGTGACGGCTTTGGGGTTCAGCCCACTCCCGGAGGAG 365
QY      81 ArgLeuAlaValLeuLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db      366 CAGCTGGGTGCTGCTGGCAGGATGCACACACTGCTCCAGCAGGCTCCTCCGTCTGCC 425
QY      101 ProAlaSerLeuLysProAspAlaPheGlnGluGlnLeuGlnGlyIleProGln 120
Db      426 CCCACCAGCCTGAAGCCTGACACCTTCAGGGACAGCTCCAGGAGCTTCGATCCCCCAA 485
QY      121 AspleulleGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db      486 GACCTGGTGGGGACTTGGCCAGCGGTGATTTATTTGGGAGCCAGCGGCCCTCTGATTCT 545
QY      141 ValAlaGlnGlnGlnGlySerSerLeuProHisValSerTyrPheArgTrpArgValAsp 160
Db      546 GTGCCCCAGCAGCAGGGGGCTGGCTGCCCATGTTCTGACTTTTCGGTGGCGGGTGGAT 605
QY      161 ValAlaLeuSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db      606 GTAGCAATCTCCACAGCTGGCTGCTGCTGCCCTGAGCCGAGCGCTCTGTATGCAGCTG 665
QY      181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
Db      666 AAGCTTCAGATGGTCAGCATACCGCTTTGAGTCCCCACAGCCAAAGTTCCAGGAGCTG 725
QY      201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGluLysCysGluArg 220
Db      726 CGGTACAGCTGGCCCTGGTCTCTAAAGAGATGCGAGATCTGGAGAAGAGGTGTGAGCGC 785
QY      221 LysLeuGlnAsp 224
Db      786 AGACTGCAGGAC 797
RESULT 7
LOCUS      BC003055      1124 bp      mRNA      linear      PRI 12-JUL-2001
DEFINITION      Homo sapiens, HT002 protein; hypertension-related calcium-regulated
                  gene, clone MGC:947 IMAGE:3544106, mRNA, complete cds.
ACCESSION      BC003055
VERSION      BC003055.1      GI:12804384
KEYWORDS      MGC.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1124)
Strausberg, R.
Direct Submission
Submitted (06-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
```

Tissue Procurement: ATCC/DCTD/DRP
 cDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywnski, Reza Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Nees, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 8 Row: A Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10434819.

FEATURES

source
 1. 1124
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /db_xref="taxon:9606"
 /clone="MGC:947 IMAGE:3544106"
 /issue_type="Skin, melanotic melanoma."
 /clone_lib="NIH_MGC_20"
 /lab_host="DH10B-R"
 /note="Vector: POTB7"
 205. 879
 /codon_start=1
 /product="HT02 protein; hypertension-related
 calcium-regulated gene"
 /protein_id="AAH03055.1"
 /db_xref="GI:12804385"
 /translation="MSAVGAATPYLHPHGDHSGRVFLGQLPPEVAAARLLGDD
 RSTPRKLKFFVSLGDEDCREAVORLGVSNLPEECGALLAGMHTLLODALPPT
 SLKPTFRDLOELCTIPDLDVGLASVFGSORPLDSVAQCGAMLPVHADFWRVD
 VAISTALARSLOPSVLMQKLKSDSAYRFEVPAKQELRYVALVLEKMADEKRC
 ERRLOD"

CDS

BASE COUNT 243 a 304 c 349 g 228 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2 916-71 Length: 1124
 Score: 913.00 Matches: 180
 Percent Similarity: 90.18% Conservative: 22
 Best Local Similarity: 80.36% Mismatches: 22
 Query Match: 82.10% Indels: 0
 DB: 9 Gaps: 0

US-09-904-568-2 (1-224) x BC003055 (1-1124)

QY 1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
 Db 205 ATGTCTGCTGTTGGGGCTGCACTCATACCTGCATCTCGTGATAGTCACAGTGGC 264
 QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeu 40
 Db 265 CGAGTGAAGTCTTTTGGGGCCCACTTCTCCAGAGGTGGCAGCAATGCCCGGCTACTA 324
 QY 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValAlaGlyAlaLeuHis 60
 Db 325 GGGGACCTAGACAGGACGCTTGGAAAGTTGCTGAAAGTTTGGCTAGCAGCCTGAG 384
 QY 61 GlyLysAspCysArgGluAlaValGluGlnLeuGlyAlaSerAlaAsnLeuSerGluGlu 80
 Db 385 GGGGAGGAGCTCCGAGAGGCTGTGACGCTTGGGGGTACAGCCCAACTGCGGAGAG 444
 QY 81 ArgLeuAlaValAlaLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db 445 CAGCTGGGTGCTCTGCTGCGCAGCATGCACTGCTCCAGAGGCCCTCTGCTGCC 504

QY 101 ProAlaSerLeuLysProAspAlaPheGlnGlnGluLeuGlnGluLeuGlyAlaProGln 120
 Db 505 CCCACCGCCTGAAACCTGACACACTTTCAGGAGCAGCAGCTCCAGAGCTGATCCCA 564
 QY 121 AspLeuIleGlyAspLeuAlaSerLeuAlaPheGlnGlnGluLeuGlnGluLeuAspSer 140
 Db 565 GACCTGCTCGGAGACTTGGCCGAGCTGTATTTGGAGCCAGCCGCGCCCTCTGATTTCT 624
 QY 141 ValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
 Db 625 GTGGCCAGCAGCAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 684
 QY 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValAlaLeuGlnLeu 180
 Db 685 GTAGCAATCTCCACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
 QY 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGlnLeu 200
 Db 745 AAGCTTCAAGTGGGTCAAGCATACCTGTTGAGGTCTCCACAGCCAGCAAGTTCCAGAGCTG 804
 QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGlnGlnGlnGlnGlnGln 220
 Db 805 CGGTACACCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 864
 QY 221 LysLeuGlnAsp 224
 Db 865 AGACTGCAGGAC 876

RESULT 8

AF290195 1355 bp mRNA linear PRI 10-OCT-2000
 AF290195 Homo sapiens hypertension-related calcium-regulated gene mRNA,
 complete cde.
 AF290195
 AF290195.1 GI:9964286

REFERENCE

1 (bases 1 to 1355)
 Solban,N., Jia,H.-P., Richard,S., Tremblay,S., Devlin,A.M.,
 Peng,J., Gossard,F., Guo,D.-F., Morel,G., Hamet,P., Lewanczuk,R.
 and Tremblay,J.
 HCAR2, a novel calcium-regulated gene coding for a nuclear protein,
 is potentially involved in the regulation of cell proliferation
 J. Biol. Chem. 275 (41), 32234-32243 (2000)

REFERENCE

2 (bases 1 to 1355)
 Tremblay,J., Solban,N., Lewanczuk,R., Jia,H.-P. and Hamet,P.
 Direct Submission
 Submitted (26-Jul-2000) CHUM Research Center, University of
 Montreal, St-Urbain, Montreal, Qc H2W 1T8, Canada
 Location/Qualifiers
 1. 1355

FEATURES

source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /issue_type="parathyroid"
 295. 969
 /note="hypertension-related calcium-regulated gene CAR2"
 /codon_start=1
 /product="unknown"
 /protein_id="AA09915.1"
 /db_xref="GI:9964287"
 /translation="MSAVGAATPYLHPHGDHSGRVFLGQLPPEVAAARLLGDD
 RSTPRKLKFFVSLGDEDCREAVORLGVSNLPEECGALLAGMHTLLODALPPT
 SLKPTFRDLOELCTIPDLDVGLASVFGSORPLDSVAQCGAMLPVHADFWRVD
 VAISTALARSLOPSVLMQKLKSDSAYRFEVPAKQELRYVALVLEKMADEKRC
 ERRLOD"

CDS

BASE COUNT 273 a 394 c 394 g 293 t 1 others
 ORIGIN

Alignment Scores:
Pred. No.: 3.68e-71 Length: 1355
Score: 913.00 Matches: 180
Percent Similarity: 90.18% Conservative: 22
Best Local Similarity: 80.36% Mismatches: 22
Query Match: 82.10% Indels: 0
DB: 9 Gaps: 0

US-09-904-568-2 (1-224) x AF290195 (1-1355)

QY 1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
Db 295 ATGTCTGTCTGGGGCTGCAACTCATACCTGCATCATCTCTGGTGAAGTACAGAGTGC 354
QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40
Db 355 CGAGTGAAGTTCTTGGGGCCCGAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 414
QY 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db 415 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGTCCAGCAGCCTGCAG 474
QY 61 GlyLysAspCysArgGluAlaValGluGlnLeuGlyAlaSerAlaAsnLeuSerGluGlu 80
Db 475 GGGGAGGAGCTGCCGAGAGCTGTGCAGCGCTTGGGGTTCAGGCCCACTGCCGGAGGAG 534
QY 81 ArgLeuAlaValLeuLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 535 CAGCTGGTGCCTGTGGCAGGATGCACACTGCTCCAGCAGGCCCTCCGTCCTGCC 594
QY 101 ProAlaSerLeuLysProAspAlaPheGlnGluGluGlnLeuGlyLeuProGln 120
Db 595 CCCACAGCCTGAAGCCTGACACTTCAGGAGCAGCTCCAGGAGCTCTGCATCCCCCAA 654
QY 121 AspleuileGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db 655 GACCTGTGGGACCTTGGCAGCGTGTATTTGGGAGCCAGCGGCCCTCTCTGATTCT 714
QY 141 ValAlaGlnGlnGlnGlySerSerLeuProHisValSerTyrPheArgTrpArgValAsp 160
Db 715 GTGGCCAGCAGCAGGGGCTGCTGCCGATGTTGCTGACTTTCGGTGGCGGGTGGAT 774
QY 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 775 GTAGCAATCTCCACAGTGCCTGCTGCTCCCTGAGCGAGCGCTCTGATGACGCTG 834
QY 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
Db 835 AAGCTTTTCAGATGGTTCAGCATACCGCTTTGAGGTCCTCCACAGCAAGTTCCAGGAGCTG 894
QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGluLysLysCysGluArg 220
Db 895 CGGTACAGCGTGGCCCTGCTCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGC 954
QY 221 LysLeuGlnAsp 224
Db 955 AGACTGCAGGAC 966

RESULT 9
LOCUS AK023070 3112 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ13008 fis, clone NT2RP3000456.
ACCESSION AK023070
VERSION AK023070.1 GI:10434819
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA,
clone_lib:NT2RP3 clone:NT2RP3000456.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES
source

CDS

BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 1.05e-70 Length: 3112
Score: 913.00 Matches: 180
Percent Similarity: 90.18% Conservative: 22
Best Local Similarity: 80.36% Mismatches: 22
Query Match: 82.10% Indels: 0
DB: 9 Gaps: 0

US-09-904-568-2 (1-224) x AK023070 (1-3112)

QY 1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
Db 234 ATGTCTGTCTGGGGCTGCAACTCATACCTGCATCATCTCTGGTGAAGTACAGAGTGC 293
QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40
Db 294 CGAGTGAAGTTCTTGGGGCCCGAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 353
QY 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db 354 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGTCCAGCAGCCTGCAG 413
QY 61 GlyLysAspCysArgGluAlaValGluGlnLeuGlyAlaSerAlaAsnLeuSerGluGlu 80
Db 414 GGGGAGGAGCTGCCGAGAGCTGTGAGCGCTCTTGGGGTTCAGGCCCACTGCCGGAGGAG 473
QY 81 ArgLeuAlaValLeuLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 474 CAGCTGGTGCCTGTGGCAGGATGCACACTGCTCCAGCAGGCCCTCCGTCCTGCC 533

Nishikawa T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tani,H., Kimata,M., Watanabe M., Hiraoka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,
Masuo,Y. and Kanehori,K.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3112)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

Location/Qualifiers

1..3112
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3000456"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP3"
/notes="Cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
234..908
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB14389.1"
/db_xref="GI:10434820"
/translation="MSAVGAATPYLHPGDSHSGRVFLCAQLPPEVAMARLLGLDLD
RSTFRKLLKEVSSLOGEDCREAVQRLGVSANLPEEQGLALLAGMTLLQOALRLPPT
SLKPDTPRDLOQLCI PQDLVGLASVFGSORPLDLSVAQQGQGLPHVADPRWRVD
VALTSALARSLOPSVLMLKLSDGSAYRFEVPTAKFQELRYVALVLKEMADLEKRC
ERRIQD"

BASE COUNT 665 a 904 c 909 g 634 t

variation /note="T substituted in clone: RP4-718C10"
/replace="C"
18051
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
18282
variation /note="A substituted in clone: RP4-718C10"
/replace="G"
19734.19737
/note="deleted in clone: KOX18a60b02.t1, deleted in
clone: RPI-316H24, deleted in clone: RPI-291P5, deleted
in clone: RP5-1056B24"
/replace=""
20793
variation /note="C substituted in clone: RP4-718C10"
/replace="A"
22023
variation /note="deleted in clone: RP4-626A24, deleted in clone:
RPI-316H24, deleted in clone: KOX18a55C01.t1, deleted in
clone: RPI-291P5, deleted in clone: RP5-1056B24, deleted
in clone: KOXa50d08.t1"
/replace=""
23997
variation /note="T substituted in clone: RP4-718C10"
/replace="C"
24217
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
24370
variation /note="T substituted in clone: RP4-718C10"
/replace="C"
29749
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
30083.30087
variation /note="GAGCA substituted in clone: RP4-718C10"
/replace="TGTTTC"
30702
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
34623
variation /note="deleted in clone: RP5-1056B24"
/replace=""
34673
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
34917
variation /note="C substituted in clone: RP4-718C10"
/replace="T"
36518
variation /note="C substituted in clone: RP4-718C10"
/replace="A"
36829
variation /note="G substituted in clone: RP4-718C10"
/replace="A"
37705
variation /note="A substituted in clone: RP4-718C10"
/replace="C"
38091
variation /note="A substituted in clone: RP4-718C10"
/replace="G"
38111
variation /note="G substituted in clone: RP4-718C10"
/replace="C"
38134
variation /note="T substituted in clone: RP4-718C10"
/replace="C"
38251
variation /note="A substituted in clone: RP4-718C10"
/replace="G"
38372
variation /note="A substituted in clone: RP4-718C10"
/replace="G"

variation 38546
/note="C substituted in clone: RP4-718C10"
/replace="G"
39256
variation /note="C substituted in clone: RP4-718C10"
/replace="T"
40089
variation /note="A substituted in clone: RP4-718C10"
/replace="G"
40506
variation /note="A substituted in clone: RP4-718C10"
Alignment Scores:
Pred. No.: 5.28e-69 Length: 69587
Score: 913.00 Matches: 180
Percent Similarity: 90.18% Conservatives: 22
Best Local Similarity: 80.36% Mismatches: 0
Query Match: 82.10% Indels: 0
DB: 9 Gaps: 0
US-09-904-568-2 (1-224) x AF146367 (1-69587)
Qy 1 MetSerAlaLeuGlyAlaAlaProTyrLeuHisHisProAlaAspSerHisSerGly 20
Db 36814 ATGCTGCTGTGGGGCTGCAACTCCATACCTGATCATCTCGTGTAGTACAGTGGC 36873
Qy 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeu 40
Db 36874 CGAGTGAGTTCTTTGGGGGGCCAGCTTCTCCAGAGTGGCAGCAATGCCCGGTACTA 36933
Qy 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db 36934 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTGCTGAAGTTTGGTTCAGCAGCTGCAG 36993
Qy 61 GlyLysAspCysArgGluAlaValGlnLeuGlyAlaSerAlaAsnLeuSerGluGln 80
Db 36994 GGGGAGGACTGCCGAGAGCTGTGCAGCGCTCTTGGGGTTCAGCGCAACCTGCCGAGGAG 37053
Qy 81 ArgLeuAlaValLeuLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 37054 CAGCTGGGTGCCCTGCTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCGCTCTGCC 37113
Qy 101 ProAlaSerLeuLysProAspAlaPheGlnGluLeuGlnGluLeuGlyLeProGln 120
Db 37114 CCCACCAGCTGAAGCCTGACACCTTCAGGAGCAGCTCCAGGAGCTCTGCATCCCCAA 37173
Qy 121 AspLeuIleGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db 37174 GACCTGGTCCGGGACTTGGCCAGCGTGTATTTGGGAGCCAGCGGCCCTCTCTGATTCT 37233
Qy 141 ValAlaGlnGlnGlySerSerLeuProHisValSerTyrPheArgTyrArgValAsp 160
Db 37234 GTGGCCACAGCAGCGGGGCTGGCTGCCGATGTTGCTGACTTTCGGTGGCGGTGGAT 37293
Qy 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 37294 GTAGCAATCTCCACCAGTGCCTCGCTCCCTGCAGCGGAGCGTCTCTGATGCAGCTG 37353
Qy 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
Db 37354 AAGCTTTCAGATGGGTACGATACCGCTTTGAGGTCCCCACAGCCAAAGTTCCAGAGCTG 37413
Qy 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGluLysCysGluArg 220
Db 37414 CGGTACAGGTGGCCCTGGTCTCTAAAGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGC 37473
Qy 221 LysLeuGlnAsp 224
Db 37474 ACAGCTGAGGAC 37485
RESULT 11
AC130271
LOCUS AC130271 127805 bp DNA linear PRI 09-AUG-2002

Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 156503)
Birtner, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferrera, P., Fitzgerald, M., Gage, D., Galagan, J., Gardina, S., Gord, S., Graham, L., Grant-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Kataras, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., PhunKhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:1098964.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1038_A.13
Center clone name: 111556
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149861 bases at least Q40
Consensus quality: 153189 bases at least Q30
Consensus quality: 154502 bases at least Q20
Insert size: 15200; agarose-fp
Insert size: 15203; sum-of-contigs
Quality coverage: 7.6 in Q20 bases; agarose-fp
Quality coverage: 7.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 14239: contig of 14239 bp in length
* 14240 14339: gap of 100 bp
* 14340 15022: contig of 683 bp in length
* 15023 15122: gap of 100 bp
* 15123 15656: contig of 534 bp in length
* 15657 15756: gap of 100 bp
* 15757 15924: contig of 168 bp in length
* 15925 16024: gap of 100 bp
* 16025 16690: contig of 666 bp in length
* 16691 16790: gap of 100 bp
* 16791 17526: contig of 736 bp in length
* 17527 17626: gap of 100 bp

FEATURES
source
* 17627 18390: contig of 764 bp in length
* 18391 18490: gap of 100 bp
* 18491 19990: contig of 1500 bp in length
* 19991 20090: gap of 100 bp
* 20091 69060: contig of 48970 bp in length
* 69061 69160: gap of 100 bp
* 69161 74866: contig of 5706 bp in length
* 74867 74966: gap of 100 bp
* 74967 90415: contig of 15450 bp in length
* 90417 90516: gap of 100 bp
* 90517 116021: contig of 25505 bp in length
* 116022 116121: gap of 100 bp
* 116122 155094: contig of 38973 bp in length
* 155095 155194: gap of 100 bp
* 155195 156503: contig of 1309 bp in length.
Location/Qualifiers
1..156503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP5-1038A13"
/clone_lib="RPCT Human PAC library 5"
1..14239
/note="assembly_fragment
clone_end:SP6
vector_side:left"
14340..15022
/note="assembly_fragment"
15123..15656
/note="assembly_fragment"
15757..15924
/note="assembly_fragment"
16025..16690
/note="assembly_fragment"
16791..17526
/note="assembly_fragment"
17627..18390
/note="assembly_fragment"
18491..19990
/note="assembly_fragment"
20091..69060
/note="assembly_fragment"
69161..74866
/note="assembly_fragment"
74967..90416
/note="assembly_fragment"
90517..116021
/note="assembly_fragment"
116122..155094
/note="assembly_fragment"
155195..156503
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 42139 a 37810 c 36308 g 38944 t 1302 others
ORIGIN
Alignment Scores:
Pred. No.: 1,47e-68 Length: 156503
Score: 913.00 Matches: 180
Percent Similarity: 90.18% Conservative: 22
Best Local Similarity: 80.36% Mismatches: 22
Query Match: 82.10% Indels: 0
Gaps: 0
US-09-904-568-2 (1-224) x AC084347 (1-156503)
QY 1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisIshIsPAlaApeSerHisSerGly 20
Db 131733 ATGCTGCTGCGGGGCGTGCACATCCATACCTGATATCTCTGTATATCATCAGAGGCC 131792
QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValThrAlaValAlaGlnLeuLeu 40

```
|||||
Db 131793 CGAGTGCAGTTCTTGGGGCCCGCAGCTTCTCCAGAGGTGGCAGCAATGGCCCGGTACTTA 131852
Qy 41 LysAspLeuAspArgSerThrPheArgLysLeuLeuLysLeuValValGlyAlaLeuHis 60
Db 131853 GGGGACCTAGACGAGGACGTTTCAAGATTGCTGAAGTTGTGGTTCAGCAGCTTCGAG 131912
Qy 61 GlyLysAspCysArgGluAlaValGluGlnLeuGlyAlaSerAlaAsnLeuSerGluGlu 80
Db 131913 GGGGAGGACTGCCGAGAGCTGTGCAGCGCTTTGGGTTCAGCGCCAACTGCCGGAGGAG 131972
Qy 81 ArgLeuAlaValLeuAlaGlyThrHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 131973 CAGCTGGGTGCCCTGTGTGCAGGCATGCACACTGCTCCAGCAGGCCCTCCGTCTGCC 132032
Qy 101 ProAlaSerLeuLysProAspAlaPheGlnGluLeuGlnGluLeuGlyLeuProGln 120
Db 132033 CCACAGGCTGAAGCTGACACTTTCAGGACCACTTCAGGAGCTTCGAGAGCTTCGCAATCC 132092
Qy 121 AspLeuIleGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db 132093 GACCTGGTGGGAGCTTGGCCAGCGTGGTATTGGGAGCCAGCGCCCTCTCTGATTCT 132152
Qy 141 ValAlaGlnGlnGlySerSerLeuProHisValSerTyrPheArgTrpArgValAsp 160
Db 132153 GTGGCCACGACGAGGGGCGCTGGCTGCCGCATGTTGCTGACTTTCGGTGGCGGTGGAT 132212
Qy 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 132213 GTAGCAATCTCCACCAAGTCCCTGCTGCTCCCTCCAGCCAGCGTCTGTATGAGCTG 132272
Qy 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
Db 132273 AGCTTTCAGATGGGTTCAGCATACCGCTTTGAGTCTCCACACGCCAAGTTCAGGAGCTG 132332
Qy 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaGluLeuGluLysCysGluArg 220
Db 132333 CGGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTCGAAGAAGGTTGTGAGCGC 132392
Qy 221 LysLeuGlnAsp 224
Db 132393 AGACTGCAGGAC 132404

RESULT 13
LOCUS AF216665
DEFINITION Homo sapiens chromosome 8 clone RP5-1056B24 map 8qtel, complete
sequence.
ACCESSION AF216665
VERSION AF216665.4 GI:22002122
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163457)
AUTHORS Schudy,A., Blechschmidt,K., Menzel,U., Polley,A., Reichwald,K.,
Rump,A., Schilhabel,M.B., Taudien,S., Wen,G., Koczan,D.,
Thiesen,H.-J., Merck-Rousseau,W.F. and Rosenthal,A.
TITLE Chromosome 8 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163457)
AUTHORS Schudy,A., Schilhabel,M., Baumgart,C., Menzel,U., Weber,J.,
Schattevoy,R. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 163457)
AUTHORS Lagemann,D. and Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (29-JUL-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On Jul 29, 2002 this sequence version replaced gi:14327859.
```

```
----- Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: H224
Center clone name: RP5-1056B24
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 163451 bases at least Q40
Consensus quality: 163451 bases at least Q30
Consensus quality: 163454 bases at least Q20
Quality coverage: 23.64x
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source
1.163457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8qtel, complete sequence."
/clone="RP5-1056B24"
48442..157868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP1-316H24"
82810..152396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP4-718C10"
514
/note="G substituted in clone: CTD-2300118"
/replace="A"
1355
/note="G substituted in clone: CTD-2300118"
/replace="T"
1432
/note="G substituted in clone: RP5-1047A19"
/replace="A"
2574
/note="A substituted in clone: CTD-2300118"
/replace="G"
3147..3150
/note="deleted in clone: RP5-1109M23, deleted in clone:
RP5-1047A19, deleted in clone: RP5-1056B24, deleted in
clone: RP5-1111B24, deleted in clone: RP1-291P5, deleted
in clone: RP4-626A24, deleted in clone: RP5-1124C13"
/replace="N"
3172..3477
/note="deleted in clone: RP5-1109M23, deleted in clone:
RP5-1047A19, deleted in clone: RP5-1056B24, deleted in
clone: RP5-1111B24, deleted in clone: RP1-291P5, deleted
in clone: RP4-626A24, deleted in clone: RP5-1124C13"
variation
variation
variation
variation
variation
variation
```

variation /replace=""
3478 /note="T substituted in clone: CTD-2300118"
/replace="C"
variation /note="C substituted in clone: CTD-2300118"
4100 /replace="A"
5038 /note="A substituted in clone: CTD-2300118"
/replace="C"
variation /note="A substituted in clone: CTD-2300118"
7130 /replace="G"
7224 /note="G substituted in clone: CTD-2300118"
/replace="A"
7396 /note="deleted in clone: RPS-1056B24 , deleted in clone:
RPI-291P5 , deleted in clone: RPS-1124C13 , deleted in
clone: CTD-2300118 , deleted in clone: RPS-111B24 ,
deleted in clone: RPA-626A24"
/replace=""
variation /note="deleted in clone: RPI-291P5 , deleted in clone:
RPS-1056B24 , deleted in clone: RPS-1124C13 , deleted in
clone: RPS-1047A19 , deleted in clone: RPS-111B24"
/replace=""
variation /note="deleted in clone: CTD-2300118"
/replace=""
10171 /note="C substituted in clone: RPS-1056B24"
/replace="T"
12318 /note="G substituted in clone: CTD-2300118"
/replace="A"
12826 /note="T substituted in clone: CTD-2300118"
/replace="C"
12844 /note="G substituted in clone: CTD-2300118"
/replace="A"
12850 /note="G substituted in clone: CTD-2300118"
/replace="T"
12905 /note="A substituted in clone: CTD-2300118"
/replace="C"
15231 /note="C substituted in clone: CTD-2300118"
/replace="T"
15348 /note="T substituted in clone: CTD-2300118"
/replace="C"
15730 /note="T substituted in clone: CTD-2300118"
/replace="A"
15750 /note="A substituted in clone: CTD-2300118"
/replace="C"
17931 /note="deleted in clone: RPS-1124C13 , deleted in clone:
RPS-1056B24 , deleted in clone: RPS-111B24 , deleted in
clone: RPI-291P5 , deleted in clone: RPA-626A24 , deleted
in clone: RPS-1047A19"
/replace=""
19197 /note="deleted in clone: CTD-2300118"
/replace=""
19197 /note="deleted in clone: CTD-2300118"
/replace=""
20259 /note="deleted in clone: RPS-1047A19"
/replace=""

variation /note="G substituted in clone: RPS-1124C13"
/replace="A"
20809 /note="G substituted in clone: CTD-2300118"
/replace="A"
20914 /note="A substituted in clone: CTD-2300118"
/replace="G"
20919 /note="T substituted in clone: CTD-2300118"
/replace="C"
23659 /note="deleted in clone: RPS-1124C13 , deleted in clone:
RPS-1056B24 , deleted in clone: CTD-2300118"
/replace=""
23659 /note="deleted in clone: RPI-291P5"
/replace=""
23659 /note="G substituted in clone: RPS-111B24"
/replace="A"
23677 /note="deleted in clone: RPI-291P5 , deleted in clone:
RPS-1047A19"
/replace=""
23862 /note="deleted in clone: RPI-291P5 , deleted in clone:
RPS-1047A19"
/replace=""
23862 .24078

Alignment Scores:

Pred. No.: 1,556-68 Length: 163457
Score: 913.00 Matches: 180
Percent Similarity: 90.18% Conservative: 22
Best Local Similarity: 80.36% Mismatches: 22
Query Match: 82.10% Indels: 0
DB: 9 Gaps: 0

US-09-904-568-2 (1-224) x AP21665 (1-163457)

QY 1 MetSerAlaLeuGlyAlaAlaAlaProTyrLeuHisPheAlaAspSerHisSerGly 20
Db 119623 ATGTCTGCTGGGGGCTGCAACTCCATCTGATCTCTGGATGATGACAGTGGC 119682
QY 21 ArgValSerPheLeuGlySerGlnProSerProGluValIleAlaAlaGlnLeu 40
Db 119683 CGAGTGAATTCCTGGGGGCTGACCTTCCTCAAGGAGGAGGAGGAGGAGGAGGAG 119742
QY 41 LysAspLeuAspArgSerThrPheArgHisLeuValValGlyAlaLeuHis 60
Db 119743 GGGGACCTGAGACAGAGACCTTCAGAAAGTTGCTGAAGTTGGTCAGCAGCTGCG 119802
QY 61 GlyLysAspCysArgGluAlaValGlnLeuGlyAlaSerAlaAsnLeuSerGln 80
Db 119803 GGGGAGACTCCGAGAGGCTGTGACGCTTGGGGTCAAGCCCACTCCGAGAGGAG 119862
QY 81 ArgLeuAlaValLeuLeuAlaGlyThrHisPheLeuGlnGlnAlaLeuArgLeuPro 100
Db 119863 CAGCTGGTGCCTGCTGCGAGCATGACACACTGCTCCAGCAGGCTCTGCTGCCC 119922
QY 101 ProAlaSerLeuLysProAspAlaPheGlnGlnLeuGlnGlnGlnGlnGlnGln 120
Db 119923 CCCACAGCCTGAGGAGCTGACACTTCAGGAGCAGCAGCTCCAGAGCTCTGATCCCCA 119982
QY 121 AspleuIleGlyAspLeuAlaSerLeuAlaPheGlySerGlnArgProLeuLeuAspSer 140
Db 119983 GACCTGGTGGGGGACTGGCCAGCGTGTATTTGGAGCCAGCGGCTCTGATTTCT 120042
QY 141 ValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db 120043 GTGGCCAGCAGCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120102
QY 161 ValAlaIleSerThrSerAlaGlnSerArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 120103 GTAGCAATCTCACACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120162

Qy 181 LysLeuThrAspGlySerAlaHisArgPheGluValProIleAlaLysPheGlnGluLeu 200
 Db 120163 AAGCTTTTCAGATGGCTCAGCATACCGCTTTGAGGTCCACACAGCCAAAGTTCCAGGAGCTG 120222

Qy 201 ArgTyrSerValAlaLeuValLeuGluMetAlaGluLeuGluLysCysGluArg 220
 Db 120223 CGGTACAGCTGGCCCTGCTCTCAAGAGATGGCAGATCTCGAAGAGGGTGTGAGCGC 120282

Qy 221 LysLeuGlnAsp 224
 Db 120283 AGACTGCAGGAC 120294

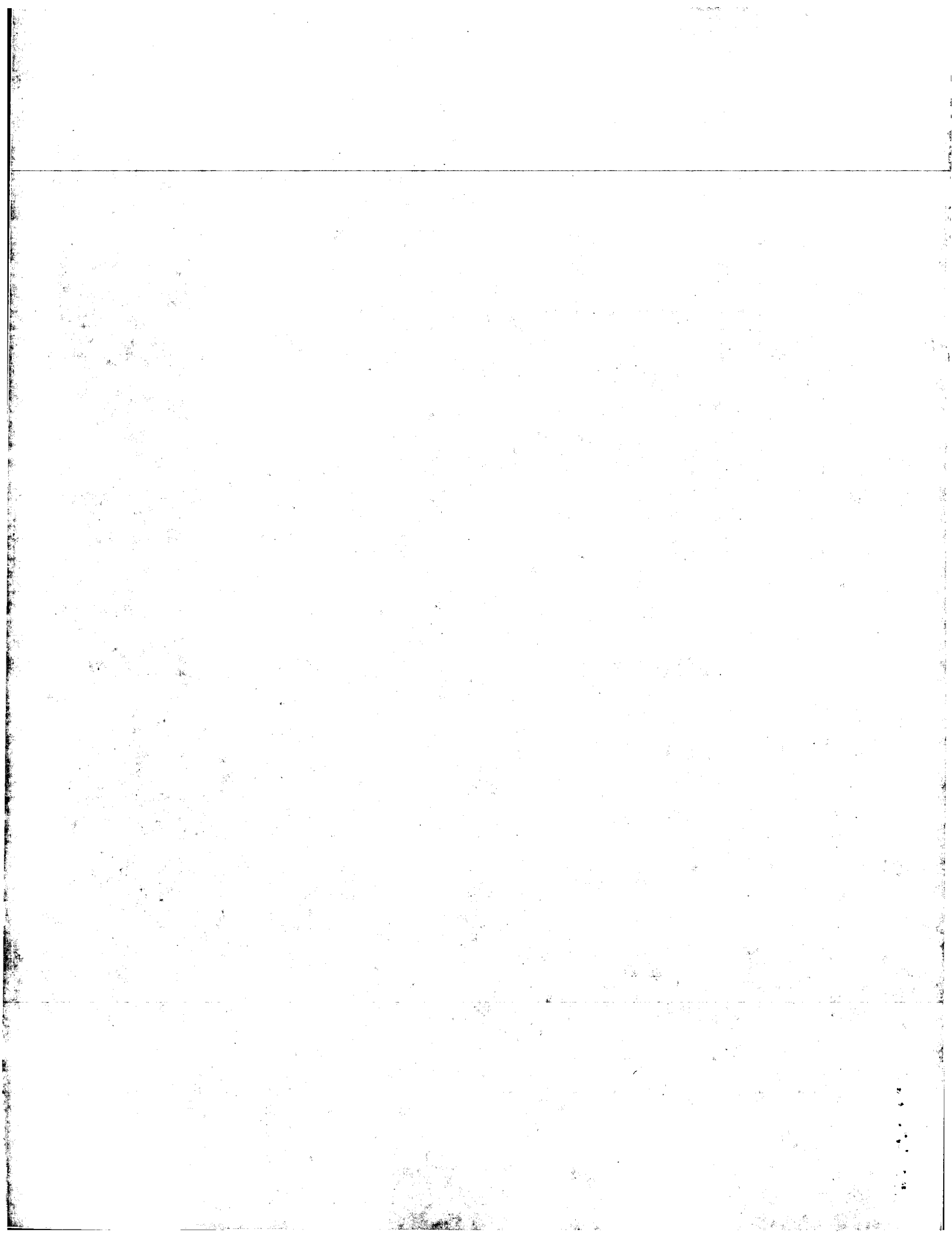
RESULT 14
 AC084856 181858 bp DNA linear HTG 22-NOV-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 8 clone CTD-3202K2 map 8, WORKING DRAFT
 AC084856
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 8, clone CTD-3202K2
 Unpublished
 2 (bases 1 to 181858)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bida,F., Boguslavskiy,L.,
 Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepell,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
 Lazarens,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
 MacDonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
 Morrow,J., Murphy,I., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Sougez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-NOV-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11037
 Center clone name: 3202 K.2
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 174342 bases at least Q40
 Consensus quality: 178412 bases at least Q30
 Consensus quality: 179957 bases at least Q20
 Insert size: 169000; agarose-fp
 Insert size: 180658; sum-of-contigs
 Quality coverage: 5.6 in Q20 bases; agarose-fp

Quality coverage: 5.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1698: contig of 1698 bp in length
 1699 1798: gap of 100 bp
 1799 4114: contig of 2316 bp in length
 4115 4214: gap of 100 bp
 4215 7476: contig of 3262 bp in length
 7477 7576: gap of 100 bp
 7577 11341: contig of 3765 bp in length
 11342 11441: gap of 100 bp
 11442 15854: contig of 4413 bp in length
 15855 15954: gap of 100 bp
 15955 25522: contig of 9568 bp in length
 25523 25622: gap of 100 bp
 25623 35216: contig of 9594 bp in length
 35217 35316: gap of 100 bp
 35317 44214: contig of 8898 bp in length
 44215 44314: gap of 100 bp
 44315 56714: contig of 12400 bp in length
 56715 56814: gap of 100 bp
 56815 68691: contig of 11877 bp in length
 68692 68791: gap of 100 bp
 68792 97288: contig of 28497 bp in length
 97289 97388: gap of 100 bp
 97389 135396: contig of 38008 bp in length
 135397 135496: gap of 100 bp
 135497 181858: contig of 46362 bp in length.

FEATURES
 source
 1. 181858
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clones="CTD-3202K2"
 /clone_lib="CITD Human BAC"
 1. 1698
 /note="assembly_fragment"
 1799. 4114
 /note="assembly_fragment"
 4215. 7476
 /note="assembly_fragment"
 7577. 11341
 /note="assembly_fragment"
 11442. 15854
 /note="assembly_fragment"
 15955. 25522
 /note="assembly_fragment"
 25623. 35216
 /note="assembly_fragment"
 35317. 44214
 /note="assembly_fragment"
 44315. 56714
 /note="assembly_fragment"
 56815. 68691
 /note="assembly_fragment"
 68792. 97288
 /note="assembly_fragment"
 97389. 135396
 /note="assembly_fragment"
 135497. 181858
 /note="assembly_fragment"
 BASE COUNT 44137 a 45562 c 45242 g 45711 t 1206 others
 ORIGIN
 Alignment Scores:



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 7, 2002, 04:21:15 ; Search time 219.5 Seconds
(without alignments)
2298.167 Million cell updates/sec

Title: US-09-904-568-4
Perfect score: 1132
Sequence: 1 MSAVCATPYLHPDHSHG.....ALVLKEMADLEKRCERLQD 224

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_p2n.model -DEV=xlp
-Q/cqn2 1/USPTO spool/US09904568/runat 25112002_143304_19913/app_query.fasta_1.782
-DB=N Geneseq 101002 -OFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -IOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=emat -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09904568 -CGCN 1.1.79 -runat 25112002_143304_19913 -NCPU=6 -ICPU=3
-NO XLPXY -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1132	100.0	1355	21	AAA75038	CDNA encoding a hu
2	1122	99.1	993	24	ABN59849	Novel human coding
3	1122	99.1	3112	22	AAH16200	Human cDNA sequenc
4	1116	98.6	995	22	AAH14347	Human cDNA sequenc
5	1084	95.8	1258	21	AAF16039	Human prostate can
6	1084	95.8	1258	24	ABL98881	Human polynucleoti
7	906	80.0	1100	19	AAV18890	Rat Hypertension r
8	906	80.0	1100	21	AAA75037	CDNA encoding a ra
9	868	76.7	684	22	AAH06228	Human cDNA clone (
10	749	66.2	704	22	AAH06388	Human cDNA clone (
11	669	59.1	2021	23	AAH88197	DNA encoding novel
12	615	54.3	701	19	AAV18891	Human Hypertension
13	461	40.7	280	22	ABA49109	Human breast cell
14	461	40.7	280	22	ABA67022	Human foetal liver
15	461	40.7	280	22	ABA34113	Probe #12579 for g
16	461	40.7	280	22	AAK15464	Human brain expres
17	461	40.7	280	22	AAK41191	Human bone marrow
18	461	40.7	280	22	AAI21949	Probe #11882 for g
19	461	40.7	280	22	AAI47232	Probe #15918 used t
20	461	40.7	280	22	AAI07640	Probe #7631 used t
21	461	40.7	280	22	ABS15202	Human genome-deriv
22	458.5	40.5	596	22	AAH11292	Human cDNA clone (
23	373.5	33.0	460	23	AAH88195	DNA encoding novel
24	314	27.7	406	24	ABN94766	Gene #1264 used to
25	289	25.5	401	22	ABA44014	Human foetal liver
26	289	25.5	401	22	ABA54469	Human breast cell
27	289	25.5	401	22	ABA24251	Probe #2717 for ge
28	289	25.5	401	22	AAK02759	Human brain expres
29	289	25.5	401	22	AAK28199	Human bone marrow
30	289	25.5	401	22	AAI12770	Probe #2703 for ge
31	289	25.5	401	22	AAI34118	Probe #2804 used t
32	289	25.5	401	22	AAI02682	Probe #2673 used t
33	289	25.5	401	24	ABS02712	Human genome-deriv
34	105.5	9.3	1302	21	AAC99872	Human secreted pro
35	105	9.3	2823	23	ABL19272	Drosophila melanog
36	105	9.3	3614	23	ABL09274	Drosophila melanog
37	102.5	9.1	2168	22	AAI18833	Human kinase (PKIN
38	97	8.6	945	23	AAH54189	Pseudomonas aerugi
39	97	8.6	8169	19	AAV26609	Actinomadura hibis
40	96.5	8.5	5169	16	AAI06001	Streptomyces wedmo
41	96.5	8.5	65140	22	AAI17184	Streptomyces nous
42	96.5	8.5	125401	22	AAI17186	Streptomyces nous
43	96	8.5	1958	24	ABA05069	Human nucleocapsid
44	95.5	8.4	2074	19	AAV32933	Human cyclin-depen
45	94.5	8.3	2074	19	AAV32932	Human cyclin-depen

ALIGNMENTS

RESULT 1
AAA75038
ID AAA75038 standard; cDNA; 1355 BP.
XX
AC AAA75038;
XX
DT 02-JAN-2001 (first entry)
XX
DE CDNA encoding a human hypertension related calcium-regulated protein.
XX
KW Hypertension related calcium-regulated protein; HCARG protein; stroke;
KW parathyroid; calcium level; gene therapy; hypertension; hyperthyroidism;
KW osteoporosis; osteopetrosis; cancer; heart failure; diabetes;
KW arteriosclerosis; gastrointestinal disease; inflammatory bowel disease;
KW asthma; ss.
XX
OS Homo sapiens.
XX

```

FH Key Location/Qualifiers
FT CDS 295..969
FT /*tag= a
FT /product= "hypertension related calcium-regulated
FT protein"
PN CA2256123-A1.
XX
XX 30-JUN-2000.
XX
XX 31-DEC-1998; 98CA-2256123.
XX
XX 31-DEC-1998; 98CA-2256123.
XX
XX (UWMO-) UNIV MONTREAL CENT RECH CENT HOSPITALIER.
XX
XX Gossard F, Lewanczuk R, Hamet P, Tremblay J;
XX
XX WPI; 2000-587793/56.
XX
XX P-PSDB; AAB08839.
XX
XX Novel hypertension related, calcium-regulated gene isolated from
XX parathyroid of a mammal, useful e.g. for treating hypertension,
XX osteoporosis, osteopetrosis, cancer, heart failure, diabetes and
XX arteriosclerosis -
XX
XX Claim 3; Fig 4; 55pp; English.
XX
XX The present sequence encodes a human hypertension related calcium-
XX regulated (HCARG) protein. The HCARG gene is isolated from the
XX parathyroid, and its expression is regulated by extracellular calcium
XX concentration. The protein is a modulator of abnormal calcium levels.
XX HCARG polynucleotides and proteins are useful in gene therapy for
XX treating a disease, disorder or abnormal physical state related to the
XX abnormal modulation of calcium such as hypertension, hyperthyroidism,
XX stroke, osteoporosis, osteopetrosis, cancer, heart failure, insulin
XX dependent and independent diabetes, arteriosclerosis, gastrointestinal
XX disease, inflammatory bowel disease or asthma.
XX
XX Sequence 1355 BP; 273 A; 395 C; 394 G; 293 T; 0 other:
XX
Alignment Scores:
Pred. NO.: 7.65e-115 Length: 1355
Score: 1132.00 Matches: 224
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-904-568-4 (1-224) x AAA75038 (1-1355)
QY 1 MetSerAlaValAlaAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
DB 295 ATGTCTGCTGTGGGGGCTGCAACCTCAATCCTGCATCCTGTGATGATCACAGTGGC 354
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProProGluValAlaAlaMetAlaArgLeuLeu 40
DB 355 CGAGTGAAGTTCTTGGGGGCCAGCTTCTCCAGAGGTGGGAGCATGTGCCCGGCTACTA 414
QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
DB 415 GGGGACCTTAGACAGAGACGACCTTCGAAAGTTGCTGAAGTTTGTGTCAGACAGCTGCAG 474
QY 61 GlyIAspAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAspLeuProGlnGlu 80
DB 475 GGGGAGACATCCCAAGACGCTGTGACGCGTCTGGGGTCACGGCCCACTTCCGAGAGAG 534
QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
DB 535 CAGCTGGGTGCTGCTGTGGAGCATGACACATGCTCCGACGAGCGCTCTGCTGCC 594
QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120

```

```

DB 595 CCCACGAGCCTGAAGCCTGACACCTTCAGGAGACAGCTCCAGAGCTGTGATCCCCCA 654
QY 121 AspLeuValAlaGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
DB 655 GACCTGTGCTGGGAACTTGGCCAGCGTGATTTGGGAGGACAGCGGCCCTCTTGAATCT 714
QY 141 ValAlaGlnGlnGlnGlyAlaThrPheLeuProHisValAlaAspPheArgTyrArgValAsp 160
DB 715 GTGGCCCAAGCAGACAGGGGCGCTGCGCATGTGCTGACTTTCGTTGGCGGCTGANT 774
QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
DB 775 GTAGCAATCTCCACAGAGCGCTGCTGCTCCCTGACAGCCGAGCGCTCATGACGCTG 834
QY 181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLeuPheGlnGluLeu 200
DB 835 AAGCTTTCAGATGGGTACGATACCGCTTTGAGTCCCAACAGCCCAAGTTTCCAGAGCTG 894
QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluArg 220
DB 895 CGGTACAGCGTGGCCCTGTCTTAAAGAGATGGCAGATCTGGAGAGAGAGCTGTAGCGC 954
QY 221 ArgLeuGlnAsp 224
DB 955 AGACTGCAGAGAC 966
RESULT 2
ABN59849
ID ABN59849 standard; cDNA; 993 BP.
XX
XX ABN59849;
AC
XX
XX 28-JUN-2002 (first entry)
DT
XX
XX Novel human coding sequence SEQ ID NO: 260.
DE
XX
XX Human, "antianaemic; vulnerary; antiinflammatory; immunomodulator;
XX antilethality; cerebroprotective; cyostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; ESR;
XX expressed sequence tag; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200222660-A2.
PN
XX
XX 21-MAR-2002.
PD
XX
XX 10-SEP-2001; 2001WO-US26015.
PF
XX
XX 11-SEP-2000; 2000US-0659671.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX WPI; 2002-292408/33.
DR
XX
XX P-PSDB; ABB97436.
PT
XX
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX
PS Claim 1; SEQ ID NO 260; 509pp; English.
XX
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These are isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haemopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX stroke and cancer, to screen for drugs, to treat inflammatory conditions
XX e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

```

```
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 993 BP; 205 A; 280 C; 299 G; 209 T; 0 other;
Alignment Scores:
Pred. No.: 6.26e-114 Length: 993
Score: 1122.00 Matches: 222
Percent Similarity: 99.55% Conservative: 1
Best Local Similarity: 99.11% Mismatches: 0
Query Match: 99.12% Indels: 0
DB: 24 Gaps: 0
US-09-904-568-4 (1-224) x ABN59849 (1-993)
QY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
DB 98 ATGCTGCTGTGGGGCTGCAACTCCATACCTGCATCATCTCTGGTGATAGTCACAGTGGC 157
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
DB 158 CGAGTGAAGTTCTTGGGGGCCCGACGCTTCTCCAGAGGTGGCAGCAATGGCCCGGTACTTA 217
QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
DB 218 GGGGACCTAGACAGGAGCAGTTTCAGAAAGTTCTGAAAGTTTGTGTGTCAGCAGCTGCAG 277
QY 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
DB 278 GGGGAGGACTGCCGAGAGGCTGTGCAGCGTCTTTGGGTGTCAGCGCAACCTGCCGAGGAG 337
QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
DB 338 CAGCTGGTGTGCTGTGGCAGCATGCACACACTGCTCCAGCAGGCGCTCCGTCTGCCCC 397
QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120
DB 398 CCACACAGCTGAAGCTTGACCTTCAGGAGCAGCTCCAGGAGCTCTGCATCCCCCAA 457
QY 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
DB 458 GACCTGTGGGACTTTGGCAGCGTGGTATTGGGAGCCAGCGGCCCTCTCTTGATTCT 517
QY 141 ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValalaAspPheArgTrpArgValAsp 160
DB 518 GTGCCCCAGCAGCAGGGGGCTGCTGCCCATGTTGCTGACTTTCGGTGGCGGGTGGAT 577
QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
DB 578 GTAGCAATCTCCACAGTGGCTGGCTGGCTCCCTGCGAGCCGAGCGTCTGTATGCAGCTG 637
QY 181 LysLeuSerAspGlySerAlaTyArgPheGluValProThrAlaLysPheGlnGluLeu 200
DB 638 AAGCTTTTCAGATGGGTGACGATACCGCTTTTGGGTGCTCCCAAGCAAGTTTCCAGGAGCTG 697
QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluArg 220
DB 698 CGGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTGGAGAGAGAGGTGTGAGCGC 757
QY 221 ArgLeuGlnAsp 224
DB 758 AGACTGCAGGAC 769
RESULT 3
AAH16200
ID AAH16200 standard; cDNA; 3112 BP.
XX
AC AAH16200;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15002.
XX
```

```
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPT; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 15002; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH24446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 3112 BP; 665 A; 904 C; 909 G; 634 T; 0 other;
Alignment Scores:
Pred. No.: 3.09e-113 Length: 3112
Score: 1122.00 Matches: 222
Percent Similarity: 99.55% Conservative: 1
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 99.12% Indels: 0
DB: 22 Gaps: 0
US-09-904-568-4 (1-224) x AAH16200 (1-3112)
QY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
DB 234 ATGCTGCTGTGGGGCTGCAACTCCATACCTGCATCATCTCTGGTGATAGTCACAGTGGC 293
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
DB 294 CGAGTGAAGTTCTTGGGGGCCCGACGCTTCTCCAGAGGTGGCAGCAATGGCCCGGTACTTA 353
```


Qy 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 593 GTAGCAATCTCCACAGTGCCTGGCTGGTCCCTGCGAGCCGAGCGTCTGATGAGCTG 652
Qy 181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLeu 200
Db 653 AAGCTTTTCAGATGGGTGAGCATACCGCTTTGAGGTCCCCACAGCCCAAGTTCAGGAGCTG 712
Qy 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluArg 220
Db 713 CGGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGC 772
Qy 221 ArgLeuGlnAsp 224
Db 773 AGACTGCAGGAC 784
RESULT 5
ID AAF16039 standard; cDNA; 1258 BP.
XX AAF16039;
AC AAF16039;
XX
DT 13-MAR-2001 (first entry)
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:474.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardiacactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587513/55.
DR P-PSDB; AAB56836..
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 946-947; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardiacactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.

SQ Sequence 1258 BP; 269 A; 347 C; 391 G; 243 T; 8 other;

Alignment Scores:

Pred. No.: 1.34e-109 Length: 1258
Score: 1084.00 Matches: 215
Percent Similarity: 96.88% Conservative: 2
Best Local Similarity: 95.98% Mismatches: 7
Query Match: 95.76% Indels: 0
DB: 21 Gaps: 0

US-09-904-568-4 (1-224) x AAF16039 (1-1258)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
Db 330 ATGTCGTCTGTGGGRCCTGCAACTCCATACCTGTCATCATCTCTGGTGTAGTACACAGTGC 389
Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db 390 CGAGTGAGTTCTTTGGGGGCCCGAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 449
Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db 450 GGGGACCTTAGACAKGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGTGTCAGCAGCTGCAG 509
Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db 510 GGGGAGGACTGCCGAGAGNTGTCGAGCGTCTTTGGGTTCAGCGCCAACTGCCGAGGAG 569
Qy 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 570 CAGCTGGGTGCTCTCTGGCAGGCATGACACACTGCTCCAGCAGGCCCTCGCTCTGCC 629
Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGlnLeuCysileProGln 120
Db 630 CCCACGAGCTTGAGCCTGACACCTTCAGGACACAGCTCCAGGAGCTCTGCATCCCCCA 689
Qy 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
Db 690 GACCTGGTGGGGACTTGGCCAGCGTGGTATTGGGNAGCCAGCGGCCCTCTCTGATTCT 749
Qy 141 ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp 160
Db 750 GTGGCCACGACGACGAGGGGCCCTGGCTGGCGCATGTGTGCTGACTTTGGGTGGGGTGGAT 809
Qy 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 810 GTAGCAATCTCCACAGTGCCTGGCTCGCTCCCTGCAGCCGAGCGCTCTGATGCAGCTG 869
Qy 181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLeu 200
Db 870 AAGCTTTTCAGATGGGTGAGCATACCGCTTTGAGGTCCCCCAGCCCAAGTTCAGGAGCTG 929
Qy 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuLysArgCysGluArg 220
Db 930 CGGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGC 989
Qy 221 ArgLeuGlnAsp 224
Db 990 AGACTGCAGGAC 1001
RESULT 6
ID ABL89881 standard; cDNA; 1258 BP.
XX ABL89881;
AC ABL89881;
XX
DT 24-MAY-2002 (first entry)
XX Human polynucleotide SEQ ID NO 443.
DE
KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW

KM cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 XX Homo sapiens.
 XX WO200190304-A2.
 PN 29-NOV-2001.
 XX
 XX 18-MAY-2001; 2001WO-US16450.
 XX
 XX 19-MAY-2000; 2000US-205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI, 2002-122018/16.
 XX P-PsDB: ABB89472.
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 XX
 PS Claim 4; SEQ ID NO 443; 2081bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89440-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcc_sequences.
 CC
 SQ Sequence 1258 BP; 269 A; 347 C; 391 G; 243 T; 8 other;
 Alignment Scores:
 Pred. NO.: 1.34e-109 Length: 1258
 Score: 1084.00 Matches: 215
 Percent Similarity: 96.88% Conservative: 2
 Best Local Similarity: 95.98% Mismatch: 7
 Query Match: 95.76% Indels: 0
 DB: Gaps: 0
 US-09-904-568-4 (1-224) x ABL89881 (1-1258)
 QY 1 MetSerAlaValAlaAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
 Db 330 ATGCTGCTGGGGGRCCTGCAACTCCATCCTGCATCATCTGGTGATGACAGTGGC 389
 QY 21 ArgValSerPheLeuGlyAlaGlnLeuProProGluValAlaAlaMetAlaArgLeuLeu 40
 Db 390 CGAGTGAAGTTCTTGGGGGCCCAAGCTTCTCCAGAGGTGGCAGCAATGGCCCGCTACTA 449
 QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
 Db 450 GGGGACCTAGACAKAGACAGCAGTTCGAAAGTTGGTGAAGTTGGTCAGACAGCTGAG 509
 QY 61 GlyGlnAspArgArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGln 80
 Db 510 GGGGAGAGCTGGCCGAGAGNTGCTGACGCTTGGGGGTCAAGCCCACTTCCGAGAGAG 569

QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db 570 CAGCTGGGTCCTCTCTGCGAGGCATGCACACACTGCTCCACAGAGCCCTCGTGGCC 629
 QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLysLeuProGln 120
 Db 630 CCCACCAGCTGAAAGCTTCAGACACTTCAGAGGACCGAGCTCCAGAGACTCTGCATCCCCCA 689
 QY 121 AspLeuValGlyAspLeuAlaSerValAlaPheGlySerGlnArgProLeuLeuAspSer 140
 Db 690 GACCTGTGGGACTTGGCCAGCGTGTATTTGGGAGCCAGCCGCCCTCTTGATTTCT 749
 QY 141 ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp 160
 Db 750 GTGGCCCAAGCAGCAGGGGGCCCTGCGCATGTGCTGACTTTCGATGGCGGGTGAT 809
 QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
 Db 810 GTAGCAATCTCCACAGAGCTGCTGCTCCCTGCGAGCCGAGCGTCTGATGAGCTG 869
 QY 181 LysLeuSerAspGlySerAlaTyrArgPheGlnValProThrAlaLysPheGlnLeu 200
 Db 870 AAGCTTCAGATGGGTGCAATACCGCTTTGAGGTCCCAAGCCAAAGTTCCAGAGCTG 929
 QY 201 ArgTyrSerValAlaLeuValLeuLysGlnMetAlaAspLeuGlnLysArgCysGlnArg 220
 Db 930 CGGTACACAGTGGCCCTGCTCTTAAGAGAGATGGCAGATTCGAGAGAGAGGTGGAGCC 989
 QY 221 ArgLeuGlnAsp 224
 Db 990 AGACTGCAGAGAC 1001
 RESULT 7
 ID AAV18890 standard; cDNA; 1100 BP.
 AC AAV18890;
 XX
 DT 09-JUN-1998 (first entry)
 XX
 DE Rat Hypertension related calcium regulated gene (HCARG) cDNA.
 XX
 KW Hypertension related calcium regulated gene; HCARG; rat parathyroid;
 KW extracellular calcium concentration; antibody; hypertension; ss;
 KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
 KW cancer; inflammatory disease; asthma.
 XX
 OS Rattus rattus.
 XX
 XX Key Location/Qualifiers
 FH CDS 131..806
 FT /*tag= a
 FT /product= "Hypertension related calcium regulator"
 FT misc_feature 173..196
 FT /*tag= b
 FT /note= "EF-hand like motif"
 XX
 XX MO9749807-A2.
 XX
 XX 31-DEC-1997.
 XX
 XX 23-JUN-1997; 97MO-CA00439.
 XX
 XX 21-JUN-1996; 96US-0667495.
 XX
 XX GOSSE/ GOSSE F.
 PA (HAME/) HAMET P.
 PA (LEMA/) LEWANCZUK R.
 PA (TREM/) TREMBLAY J.
 XX
 XX Gossard F, Hamet P, Lewanczuk R, Tremblay J;

DR WPI; 1998-077171/07.
XX P-PSDB; AAW37723.
XX
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
XX
PS Claim 2; Pages 25-26; 46pp; English.
XX
XX This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HcARG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, cancer,
CC inflammatory disease, and asthma.
XX
SQ Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T; 0 other;

Alignment Scores:
Pred. No.: 4,478-90 Length: 1100
Score: 906.00 Matches: 178
Percent Similarity: 89.73% Conservative: 23
Best Local Similarity: 79.46% Mismatches: 23
Query Match: 80.04% Indels: 0
DB: 19 Gaps: 0

US-09-904-568-4 (1-224) x AAV18890 (1-1100)

QY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
Db ATGTCGTCTTGGGGCTGCAGCTCCATCTTGCACCATCCCGCTCACAGTCACAGTGGC 191

QY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db CGGGTCAGTTTCTGGGTTCCCGCCCTTCCAGAAAGTCAGCGCGCTGCTCAGCTCTTG 251

QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db AAGGACTTAGACAGGAGACCTTCAGAAAGTTGTTGAACCTGTAGTCGGGGCCCTGCAT 311

QY 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db GCGAAAGACTGCAGAGAAGCTGTGGAGCAACTTGGTCCAGCGCCCACTCTCAGAAGAG 371

QY 81 GlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db CGTCTGGCGCTCTGCTGGGGGCACACACACCTCTCCAGCAGGCTCTCCGGCTGCCCC 431

QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln 120
Db CCGTCTAGTCTAAGCAGCATGCTTCCAGGAGAGCTCCAGGAACCTTGCATTCCTCAG 491

QY 121 AspLeuValGlyAspLeuAlaSerValAlpPheGlySerGlnArgProLeuLeuAspSer 140
Db GATCTAATTGGAGATTGGCCAGTTTGGCATTGGGAGTCAACGCCCTCTCTCGACTCT 551

QY 141 ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAlaPheArgTrpArgValAsp 160
Db GTAGCCCAACAGCAGGAGATCCTCGCTGCCTCAGCTGTCTTACTTCCGGTGGCGGTGGAC 611

QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db GTGGCCATCTCAACAGCGCTCAGTCCCGCTCCCTGCAACCGAGTGTCTCATGCAAGCTG 671

QY 181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLeu 200
Db AAGCTCACAGATGGATCTGCACACCGCTTCCAGGTGCCCCATAGCCAAATTCAGGAGCTG 731

QY 201 ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluArg 220
Db CGGTACAGTGTAGCTTGGTCTTAAAGGAGATGGCAGAACTGGAGAAGAAGTGTGAGCGC 791

QY 221 ArgLeuGlnAsp 224
Db 792 AAACGTGCAGGAC 803

RESULT 8
AAV75037
ID AAA75037 standard; cDNA; 1100 BP.
XX
AC AAA75037;
XX
DT 02-JAN-2001 (first entry)
XX
DE cDNA encoding a rat hypertension related calcium-regulated protein.
XX
KW Hypertension related calcium-regulated protein; HcARG protein; stroke;
KW parathyroid; calcium level; gene therapy; hypertension; hyperthyroidism;
KW osteoporosis; osteopetrosis; cancer; heart failure; diabetes;
KW arteriosclerosis; gastrointestinal disease; inflammatory bowel disease;
KW asthma; ss.
XX
OS Rattus rattus.
XX
XX
FH Key Location/Qualifiers
FT CDS 132..806
FT /*tag= a
FT /product= "hypertension related calcium-regulated
FT protein"
XX
XX CA2256123-Al.
XX
XX 30-JUN-2000.
XX
XX 31-DEC-1998; 98CA-2256123.
XX
XX 31-DEC-1998; 98CA-2256123.
XX
XX (UYMO-) UNIV MONTREAL CENT RECH CENT HOSPITALIER.
XX
XX Gossard F, Lewanczuk R, Hamet P, Tremblay J;
XX
XX WPI; 2000-587793/56.
XX P-PSDB; AAB08838.
XX
XX Novel hypertension related, calcium-regulated gene isolated from
PT parathyroid of a mammal, useful e.g. for treating hypertension,
PT osteoporosis, osteopetrosis, cancer, heart failure, diabetes and
PT arteriosclerosis -
XX
PS Claim 2; Fig 3; 55pp; English.
XX
XX The present sequence encodes a rat hypertension related calcium-
CC regulated (HcARG) protein. The HcARG gene is isolated from the
CC parathyroid, and its expression is regulated by extracellular calcium
CC concentration. The protein is a modulator of abnormal calcium levels.
CC HcARG polynucleotides and proteins are useful in gene therapy for
CC treating a disease, disorder or abnormal physical state related to the
CC abnormal modulation of calcium such as hypertension, hyperthyroidism,
CC stroke, osteoporosis, osteopetrosis, cancer, heart failure, insulin
CC dependent and independent diabetes, arteriosclerosis, gastrointestinal
CC disease, inflammatory bowel disease or asthma.
XX
SQ Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T; 0 other;

Alignment Scores:
Pred. No.: 4,478-90 Length: 1100
Score: 906.00 Matches: 178
Percent Similarity: 89.73% Conservative: 23
Best Local Similarity: 79.46% Mismatches: 23
Query Match: 80.04% Indels: 0
DB: 21 Gaps: 0

US-09-904-568-4 (1-224) x AAA75037 (1-1100)

Db 413 CCCACCGCTGACCTGACATCTTCAGGACACAGCTCCAGAGCTCTGCATCCCCCAA 472
Qy 121 AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer 140
Db 473 GACCTGGTGGGGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCCCTCTTGATTCT 532
Qy 141 ValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160
Db 533 GTGGCCCAACAGCAGGGGCTGGCTGGCGCATGTGCTGACTTTCGGTGGCGGTGGAT 592
Qy 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 593 GTACA-ACTCCACAGTGGCTGGCTCCCTGCANCCAGCGCTCTGTATGCA-CTG 650
Qy 181 LysLeuSerAspGlySerAlaTyrArgPhe 190
Db 651 AAATTTTCAGATGGGTCCNATNCCGCTTT 680

RESULT 10
AAH06388
ID AAH06388 standard; cDNA; 704 BP.
XX
AC AAH06388;
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:3223.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW Homo sapiens.
OS
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 3223; 2537pp + CD ROM; English.
XX

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 704 BP; 117 A; 201 C; 247 G; 136 T; 3 other;

Alignment Scores: 4,71e-73 Length: 704
Pred. No.: 749,00 Matches: 151
Score: 96.82% Conservative: 1
Percent Similarity: 96.82% Mismatches: 5
Best Local Similarity: 96.18% Indels: 1
Query Match: 66.17% Gaps: 0
DB: 22

US-09-904-568-4 (1-224) x AAH06388 (1-704)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
Db 234 ATGCTGCTGTGGGGCTGCAACTCCATACCTGATCATCTCTGGTGATGATCAGATGCG 293
Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db 294 CGAGTGAGTTTCTTGGGGGCCAGCTTCTCCAGAGGTGGCAGCAATGCCCGGCTACTA 353
Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLysPheValValSerSerLeuGln 60
Db 354 GGGGACCTAGACAGGACGACGTTTCAGAAAGTTGCTGAAGTTTGGTTCAGCAGCTGCAG 413
Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db 414 GGGGAGGAGTCCCGAGAGGCTGTGCAGCGTCTTGGGGTCAGGCCCACTGCCGGAGGAG 473
Qy 81 GlnLeuGlyAlaLeuAlaGlyMethHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 474 CAGCTGGGTGGCTCTGGCAGGCATGCACACACTGCTCCAGCAGGCTCTCGCTCGCCC 533
Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120
Db 534 CCCACGAGCTTGAAGCTGACACCTTCAGGGACCACTTCAGGAGCTCTGCAATGCCCAA 593
Qy 121 AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer 140
Db 594 GACCTGGTGGGGACTTGGCCANCGTGGTATTTGGAGCCAGGCCCTT-CTTGATTCT 652
Qy 141 ValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 157
Db 653 GTGGCCCAACAGCAGGGGCTGTGCCNCACTGTGCTGACTTTCGGTGG 703

RESULT 11
AAS88197
ID AAS88197 standard; cDNA; 2021 BP.
XX
AC AAS88197;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24001.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
XX 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmenac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR P-PSDB; ABG24010.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID No 24001; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIGO
 CC at ftp.wigo.int/pub/published_pcr_sequences.
 CC
 SQ Sequence 2021 BP; 440 A; 530 C; 506 G; 478 T; 67 other;
 SQ
 Alignment Scores:
 Pred. No.: 1,336-63 Length: 2021
 Score: 669.00 Matches: 194
 Percent Similarity: 81.56% Conservative: 5
 Best Local Similarity: 79.51% Mismatches: 19
 Query Match: 59.10% Indels: 26
 DB: 23 Gaps: 6
 US-09-904-568-4 (1-224) x AAS86197 (1-2021)
 QY 7 AAlaThrProlyr--LeuHISHisPro--GlyAspSerHisSer-GlyArgValSerPhe 24
 DB 805 GCCACTCCCTTAACCTTGGCATCATCTTGGGNGATGATGACAGTGGCCGAGTGGATTTC 864
 QY 25 LeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAsp 44
 DB 865 TTGGGGGCCCAAGCTTCTCCAGAGGTGCAGCAATGCGCCGCTACTAGGGGACCTAGAC 924
 QY 45 ArgSerThrPheArgGlyLeu-LeuLysPheValValSerSerLeuGlnGlyAspCys 64
 DB 925 AGGAGACAGCTTCAGAAAGTGTCTGAAGTTGTGGTGCAGAGCTGCAGGGGGAGGATGG 984
 QY 64 SARASpGlyVal--GlnArgLeuGlyValSerAlaAsnLeuProGluGlnLeuGly 83
 DB 985 CCGAGAGGCTCTCCAGCGCTTGTGGGTGAGCCCAACCTGCGGAGAGACACCTGGGT 1044
 QY 84 AlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuProThrSer 103
 DB 1045 GCCCTGCTGGAGGCAATGCAACACTGCTCCAGCAGGCGCTCTGCCCCCACCAGC 1104
 QY 104 LeuLysProAspThrPheArgAsp-GlnLeuGlnGlnLeuCysIleProGlnAspLeuVal 123

DB 1105 CTGAAGCTTGACACCTTCAGGGAGCCAGAGCTTCATCCCCCAAGACCTGCT 1164
 QY 123 lGlyAsp-LeuAlaSerValValPheGlySer--GlnArgProLeuLeuAspSerVala 142
 DB 1165 CGGGGACTTTGGCCAGCGGTGATTTGGAGGCCAAGCGGCCCTCTCTTGTCTGTGG 1224
 QY 142 laGlnGln--GlnGlyAlaTrpLeuProHisValAla-AspPheArg-TTrpArgVala 160
 DB 1225 CCCACAGACCCAGGCTGGCTGGCTGGCCGATGTTGCTTGAAGTTGGGAGTGA 1284
 QY 160 pValAlaIleSerThr-SerAla-LeuAlaArgSerLeuGln-ProSerValLeuMet-G 179
 DB 1285 TGTAACAATCTCCACCCAGTGCCTGGCTGCTGCTCCGCAAGCCGACGCTCTGATGCC 1344
 QY 179 lNLeuLys--LeuSerAspGlySerAlaTrp--ArgPheGlnVal--ProTrpAlaL 196
 DB 1345 AGCTAAAGCTTTTCAGATGGTCAAGCAATACCCGCTTTGAGGTCCCCCAGACCCCA 1404
 QY 196 ySPhe-GlnGlnLeu-ArgTyrSerValAla-LeuVal-LeuLysGluMetAlaAspLeu 214
 DB 1405 AGTCCAGAGAGCTGGCGGTACAGCGTGGCCCTGCTCTTAAGAGATGAGCATCTG 1464
 QY 215 GluLysArg-CysGlu-ArgArgLeuGlnAsp 224
 DB 1465 GAAAAAAGTTGTGAGCCGCAACTTGACAGAC 1496
 RESULT 12
 ID AAV18891 standard; cDNA: 701 BP.
 AC AAV18891;
 XX
 DT 09-JUN-1998 (first entry)
 XX
 DE Human Hypertension related calcium regulated gene (HCARG) cDNA.
 XX
 KW Hypertension related calcium regulated gene; HCARG; human parathyroid;
 KW extracellular calcium concentration; antibody; hypertension; ss;
 KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
 KW cancer; inflammatory disease; asthma.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 224..245
 FT /*tag= a
 FT /note= "BP-hand like motif"
 XX
 PN MO9749807-A2.
 XX
 PD 31-DEC-1997.
 XX
 PF 23-JUN-1997; 97MO-CA00439.
 XX
 PR 21-JUN-1996; 96US-0667495.
 XX
 PA (GOSS/) GOSSARD F.
 PA (HAME/) HAMEY P.
 PA (LENA/) LEWANCZUK R.
 PA (TREM/) TREMBLAY J.
 XX
 PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
 XX
 DR WPI: 1996-077171/07.
 XX
 PT Hypertension related calcium regulated gene - useful to develop
 PT products to treat or detect, e.g. hypertension, stroke,
 PT osteoporosis, heart failure, cancer, diabetes or asthma
 XX
 PS Claim 3; Pages 27-28; 46pp; English.
 XX
 CC This nucleic acid sequence encodes the hypertension related calcium
 CC regulated gene (HCARG), which was isolated from the human parathyroid.

CC Its expression is regulated by extracellular calcium concentration.
CC The sequence displays 75 percent homology to the rat form, with
CC conserved areas such as the EF-hand like motifs and the initiating
CC codon. An antibody against the protein, can be used to detect or
CC modulate (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.

XX Sequence 701 BP; 137 A; 204 C; 217 G; 139 T; 4 other;

Alignment Scores:
Pred. No.: 2,69e-58 Length: 701
Score: 615.00 Matches: 145
Percent Similarity: 64.63% Conservatives: 3
Best Local Similarity: 63.32% Mismatches: 10
Query Match: 54.33% Indels: 71
DB: 19 Gaps: 1

US-09-904-568-4 (1-224) x AAV18891 (1-701)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisProGlyAspSerHisSerGly 20
Db 182 ATGTCGCTTTGGGAGCTGCACTCCATCATCTGCTGATAGTCACAGTGC 241
Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaMetAlaArgLeuLeu 40
Db 242 CGAGTAGTTCCTTTGGGGCCAGCTTCCTCCAGAGTGGCAGCAATGGCCAGCTACTA 301
Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeu-LeuLysPheValValSerSerLeuG 60
Db 302 GGGACCTACAGGAGCAGCTTCAGAAAGTTGCTTGAAGTTTGGTCAGCAGCTGCA 361
Qy 60 nGlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluG 80
Db 362 GGGGGAGATTCCGAGAGTGTTCAGAGCTTTTGGGGTCAGGCCCAACCTGCCGGAGGA 421
Qy 80 uGlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnAlaLeuArgLeuPr 100
Db 422 GCAGTTGGTTCCTT----- 437
Qy 100 oProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProG 120
Db 437 ----- 437
Qy 120 nAspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSe 140
Db 437 ----- 437
Qy 140 rValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAs 160
Db 438 -----GTTGGCAGGTTTCGTTGGCGGGTGA 463
Qy 160 pValAla-IleSerThr-SerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGln 179
Db 464 TGTAGCCAATCTCCACCAGTGCCTGGCTCCCTGCGAGCGAGCGCTCTGATGACAG 523
Qy 180 Leu-LysLeuSerAspGlySerAlaTyr-ArgPheGluValProThrAlaLysPheGlnG 199
Db 524 CTGTAAGCTTTTCAGATGGGTGAGCATACCCGCTTTGAGGTCCCCACAGCCAAAGTTCCAGG 583
Qy 199 luLeuArgTyr-SerValAlaLeuValLeuLysGluMetAlaAspLeuLulysArgCysG 219
Db 584 AGCTCGGTATACAGCTGGCCCTTGCTCTTAAGGAGATGGCAGATCTGGAGNAGAGGTGTG 643
Qy 219 luArgArgLeuGlnAsp 224
Db 644 AGCCGAGACTGCAGGAC 660

RESULT 13

ABA49109

ID ABA49109 standard; DNA; 280 BP.

XX

AC ABA49109;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #7804.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes
XX
XX Claim 4; SEQ ID NO 7804; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 280 BP; 51 A; 84 C; 94 G; 51 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 6,85e-42 Length: 280
Score: 461.00 Matches: 91
Percent Similarity: 98.92% Conservatives: 1
Best Local Similarity: 97.85% Mismatches: 1
Query Match: 40.72% Indels: 0
DB: 22 Gaps: 0
US-09-904-568-4 (1-224) x ABA49109 (1-280)
Qy 11 LeuHisProGlyAspSerHisSerGlyArgValSerPheLeuGlyAlaGlnLeuPro 30
Db 1 CTGCATCATCTGCTGATAGTCACAGTGGCCGAGTTCAGTTCTTGGGGCCAGCTTCCT 60
Qy 31 ProGluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLys 50

DB 61 CCGAGGTGGCAGCAATGCGCCGGCTACTAGGGGACCTTAACAGAGGACAGGTTTCAGAAAG 120
QY 51 LeuHshSpheValValSerSerLeuGlnGlyuAspCysArgAspGlyValGlnArg 70
DB 121 TTGCTGAAGTTTGTCGTCAGCAGGCTGTCAGGGAGAGACTGCCGAGAGGCTGTGCAGCGT 180
QY 71 LeuGlyValSerAlaAsnLeuProGluGluGlnLeuGlyAlaLeuLeuAlaGlyMetHis 90
DB 181 CTGGGGTCCAGCCCAACCTGCGGAGAGACAGCTGGGTGCTGCTGTGGCAGCATGCAC 240
QY 91 ThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSer 103
DB 241 ACACTGCTCCAGCAGGCGCTCTGCTGCCCCCACCAGC 279
RESULT 14
ABA67022
ID ABA67022 standard; DNA; 280 BP.
XX
AC ABA67022;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #15327.
XX
KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
PS Claim 4; SEQ ID NO 15327; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 280 BP; 51 A; 84 C; 94 G; 51 T; 0 other;
Alignment Scores:
Pred. NO.: 6.85e-42 Length: 280
Score: 461.00 Matches: 91
Percent Similarity: 98.92% Conservative: 1
Best Local Similarity: 97.85% Mismatches: 1
Query Match: 40.72% Indels: 0
DB: 22 Gaps: 0

US-09-904-568-4 (1-224) x ABA67022 (1-280)
QY 11 LeuHshSpheValValSerSerLeuGlnGlyuAspCysArgAspGlyValGlnArg 30
DB 1 CTGCATCATCCCTGGGATGATGACAGCTGCGCCAGAGATTCTTGGGGGCCACCTTCT 60
QY 31 ProGluValAlaAlaMetAlaArgLeuLeuGlyuAspLeuAspArgSerThrPhaArgLys 50
DB 61 CCGAGGTGGCAGCAATGCGCCGGCTACTAGGGGACCTTAACAGAGGACAGGTTTCAGAAAG 120
QY 51 LeuHshSpheValValSerSerLeuGlnGlyuAspCysArgAspGlyValGlnArg 70
DB 121 TTGCTGAAGTTTGTCGTCAGCAGGCTGTCAGGGGAGAGACTGCCGAGAGGCTGTGCAGCGT 180
QY 71 LeuGlyValSerAlaAsnLeuProGluGluGlnLeuGlyAlaLeuLeuAlaGlyMetHis 90
DB 181 CTGGGGTCCAGCCCAACCTGCGGAGAGACAGCTGGGTGCTGCTGTGGCAGCATGCAC 240
QY 91 ThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSer 103
DB 241 ACACTGCTCCAGCAGGCGCTCTGCTGCCCCCACCAGC 279
RESULT 15
ABA34113
ID ABA34113 standard; DNA; 280 BP.
XX
AC ABA34113;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #12579 for gene expression analysis in human heart cell sample.
XX
KM Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID NO 12579; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.

Mon Dec 9 14:18:55 2002

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

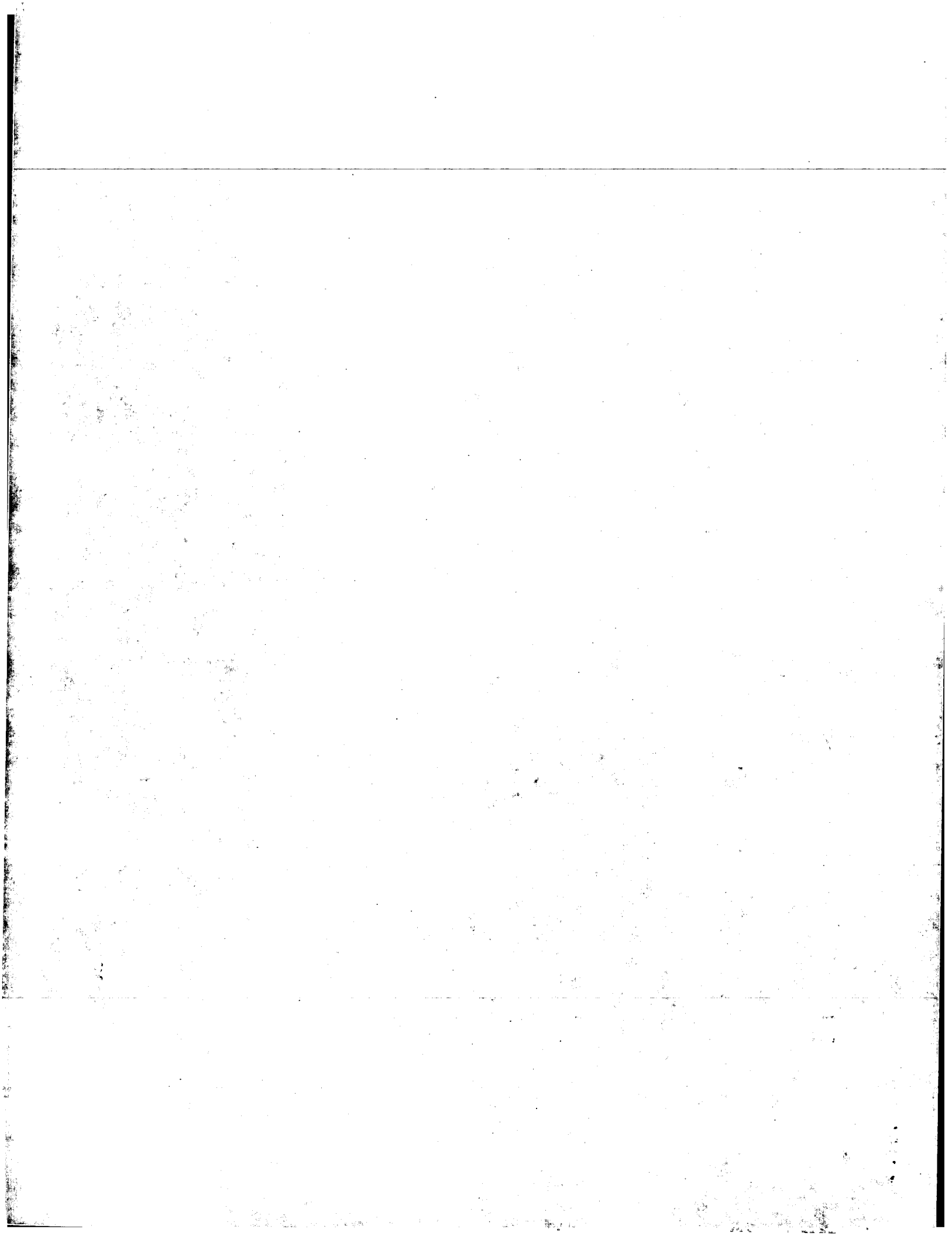
XX
SQ Sequence 280 BP; 51 A; 84 C; 94 G; 51 T; 0 other;

Alignment Scores:
Pred. No.: 6.85e-42 Length: 280
Score: 461.00 Matches: 91
Percent Similarity: 98.92% Conservative: 1
Best Local Similarity: 97.85% Mismatches: 1
Query Match: 40.72% Indels: 0
DB: 22 Gaps: 0

US-09-904-568-4 (1-224) x ABA34113 (1-280)

QY	11	LeuHisProGlyAspSerHisSerGlyArgValSerPheLeuGlyAlaGlnLeuPro	30
DB	1	CTGCATCATCTCTGGTGATAGTCACAGTGGCCGAGTGAGTTTCTTGGGGGCCCGAGCTTCCT	60
QY	31	ProGluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLys	50
DB	61	CCAGAGGTGGCAGCAATGGCCGGCTACTAGGGGACCTAGACAGGAGCACGTTTCAGAAAG	120
QY	51	LeuLeuLysPheValValSerSerLeuGlnGlyGluAspCysArgAspGlyValGlnArg	70
DB	121	TTGCTGAAGTTTGTGTGCAGCAGCCTGCGAGGGGGAGGACTGCCGAGAGGCTGTGACGCT	180
QY	71	LeuGlyValSerAlaAsnLeuProGluGlnLeuGlyAlaLeuLeuAlaGlyMethHis	90
DB	181	CTTGGGGTCAGCCCAACCTGCCGAGGAGCAGCTGGGTGCCCTGTGGCAGGCATGCAC	240
QY	91	ThrLeuLeuGlnGlnAlaLeuArgLeuProThrSer	103
DB	241	ACACTGCTCCAGCAGGCCCTTCGTCGTGCCCCGCCACCAGC	279

Search completed: December 7, 2002, 04:35:49
Job time : 223.5 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 7, 2002, 04:18:35 ; Search time 48 Seconds

(without alignments)
1431.158 Million cell updates/sec

Title: US-09-904-568-4

Perfect score: 1132

Sequence: 1 MSVGAATPYLHHFGDSHG.....ALVLKEMADLEKRCERLQD 224

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO spool/US09904568/runat_25112002_143306_19943/app_query.fasta_1.782
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pcet -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09904568 @CGN_1_1_31/runat_25112002_143306_19943 -NCPU=6 -ICPU=3
-NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/SA COMB.seq.*
2: /cgn2_6/ptodata/1/ina/SB COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95.5	8.4	2074	2	US-09-018-576-11
2	95.5	8.4	2074	3	US-09-248-137-11
3	94.5	8.3	2074	2	US-09-018-576-2
4	94.5	8.3	2074	3	US-09-248-137-2
5	93.5	8.3	751	2	US-08-822-260-2
6	89.5	7.9	1749	4	US-09-516-914-22
7	89.5	7.9	14272	4	US-09-516-914-23
8	84	7.4	3049	4	US-09-221-0178-857
9	84	7.4	9057	4	US-09-453-702B-194
10	83	7.3	4183	3	US-08-996-083-2
11	83	7.3	4183	4	US-09-429-516-2
12	82	7.2	2232	1	US-08-241-766-12

C 13	82	7.2	3120	1	US-08-491-146-1	Sequence 1, Appl
C 14	82	7.2	3120	1	US-08-241-766-11	Sequence 11, Appl
C 15	82	7.2	3120	1	US-08-234-011-1	Sequence 1, Appl
C 16	82	7.2	3120	2	US-08-701-062A-1	Sequence 1, Appl
C 17	82	7.2	3497	1	US-08-530-950-5	Sequence 5, Appl
18	82	7.2	3497	4	US-09-149-879-5	Sequence 5, Appl
19	82	7.2	3498	3	US-08-888-429A-5	Sequence 5, Appl
20	82	7.2	3553	1	US-08-530-950-7	Sequence 7, Appl
21	82	7.2	3553	4	US-09-149-879-7	Sequence 7, Appl
22	82	7.2	3554	3	US-08-888-429A-7	Sequence 7, Appl
23	82	7.2	3576	1	US-08-530-950-9	Sequence 9, Appl
24	82	7.2	3576	3	US-09-358-382-1	Sequence 1, Appl
25	82	7.2	3576	3	US-08-888-429A-9	Sequence 9, Appl
26	82	7.2	3576	4	US-09-149-879-9	Sequence 9, Appl
27	82	7.2	3587	2	US-08-874-186-91	Sequence 91, Appl
C 28	82	7.2	4403765	4	US-09-103-840A-2	Sequence 2, Appl
29	81.5	7.2	1155	2	US-08-401-068-1	Sequence 1, Appl
30	81.5	7.2	1155	2	US-08-846-338-1	Sequence 1, Appl
31	81.5	7.2	1997	2	US-08-969-630-1	Sequence 1, Appl
32	81.5	7.2	5872	3	US-08-411-768B-1	Sequence 1, Appl
33	81.5	7.2	5872	3	US-08-411-768B-6	Sequence 6, Appl
34	81	7.2	1005	2	US-08-875-062-4	Sequence 4, Appl
C 35	80	7.1	1878	4	US-09-732-025-1	Sequence 1, Appl
C 36	80	7.1	4117	4	US-09-484-970B-2	Sequence 2, Appl
37	80	7.1	28958	1	US-08-258-261B-6	Sequence 6, Appl
38	80	7.1	28958	1	US-08-456-837-6	Sequence 6, Appl
39	80	7.1	28958	1	US-08-457-342-6	Sequence 6, Appl
40	80	7.1	28958	1	US-08-457-646A-6	Sequence 6, Appl
41	80	7.1	28958	1	US-08-458-076A-6	Sequence 6, Appl
42	80	7.1	28958	1	US-08-764-233A-6	Sequence 4, Appl
43	80	7.1	28958	1	US-08-457-335A-6	Sequence 6, Appl
44	80	7.1	28958	1	US-08-729-214-6	Sequence 6, Appl
45	80	7.1	28958	3	US-09-028-934-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-09-018-576-11
; Sequence 11, Application US/09018576
; Patent No. 5968800
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,576
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2074 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

Qy 133 SerGlnArgProLeuLeuAspSerValAlaGlnGln-----GlnGly 146
Db 492 GCCAGAGGCACTAGCCAGGCAAGCTCAGAGCTACCTGCAGATGCTGCTCAAGGT 551
Qy 147 AlaTrpLeuProHisValAlaAspPheArgTrpArgValAspValAlaIleSerThrSer 166
Db 552 GTCGCCTTCGCAT-----GCCAACAC 575
Qy 167 AlaLeuAlaArgSerLeuGlnProSerValLeuMet-----GlnLeuLys 181
Db 576 ATTGTACATCGGAACCTGAACCTGCAACCTGCTCATCAGCGCTCAGCGCAGCTCAAG 635
Qy 182 LeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLys-----Phe 197
Db 636 ATAGCGGACTTTGGCTGCTGAGTCTTTCCCGAGAGCGGCGCTCTACACAC 695
Qy 198 GlnGluLeuArgTyrSerValAlaLeuValLeuLysGluMetAla----- 212
Db 696 CAGGTGGCCACCAGGCTCTGGGCTGCATCATGGGGAGCTGTGAATGGTCCGCCCTT 755
Qy 213 -----AspLeuGluLysArgCys 218
Db 756 TTCCCGGGCAAGACGATATTGAACAGCTTTGC 788

RESULT 3

US-09-018-576-2
; Sequence 2, Application US/09018576
; Patent No. 5968800
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2074 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
US-09-018-576-2

Alignment Scores:
Pred. No.: 0.0756 Length: 2074
Score: 94.50 Matches: 55
Percent Similarity: 37.85% Conservative: 40
Best Local Similarity: 21.91% Mismatches: 83
Query Match: 8.35% Indels: 73
DB: 2 Gaps: 11

US-09-904-568-4 (1-224) x US-09-018-576-2 (1-2074)
Qy 28 GlnLeuProGluValAlaAlaMetAlaArgLeuGlyAspLeuAspSerThr 47
Db 75 CAGGAACCGCCCGCTGGTGCAGCTGCCGCTGTTGTAGGCAACAGCGTGCAGGCTCAG 134
Qy 48 -----PheArgLys-----LeuLeuLysPheValValSerSerLeuGlnGly 61
Db 135 ATCAGCGTGGGGTGGAGAGAGTGGAGTTTGGAGATTTCAGGGGCACAGGGGCACAGGCC 194
Qy 62 GluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGluGln 81
Db 195 CACGACTGC-----ACGGGATGGACCACTGCTGCTGCGCCCGCTCGGGAGGGGCC 251
Qy 82 LeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 101
Db 252 CACGGCATGCTTCACAGGCCAAGCAGCTGGAGACTGGCGAGATAGTTGCCCTCAAGAAG 311
Qy 102 ThrSerLeuLys-----ProAspThrPheArgAspGln-----LeuGlnGluLeuCysile 118
Db 312 GTGGCCCTTAAGCGGCTTGGAAAGAGCGCTTCCCTAACACAGGCCCTCGGGAGATTAAAGCT 371
Qy 119 ProGlnAsp----- 121
Db 372 CTGCGAGGATGGAGGACAATCAGTATGTGTGTACAACTGAAGGCTGTTCACACACCGT 431
Qy 122 -----LeuValGlyAspLeuAlaSerValValPheGly 132
Db 432 GGAGCTTTGTGCTGGCCTTTGAGTTTCATGCTGCGGATCTGGCCGAGGTGGTGGCCAT 491
Qy 133 SerGlnArgProLeuLeuAspSerValAlaGlnGln-----GlnGly 146
Db 492 GCCAGAGGCGCACTAGCCAGGCAAGCTCAAGAGCTACCTGCAGATGCTGCTCAAGGT 551
Qy 147 AlaTrpLeuProHisValAlaAspPheArgTrpArgValAspValAlaIleSerThr 166
Db 552 GTCGCCTTCGCCAT-----GCCAACAC 575
Qy 167 AlaLeuAlaArgSerLeuGlnProSerValLeuMet-----GlnLeuLys 181
Db 576 ATTGTACATCGGACCTGAACCTGCCAACCTGCTCATCAGCGCCTCAGGCCAGCTCAAG 635
Qy 182 LeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLys-----Phe 197
Db 636 ATAGCGGACTTTGGCTGCTGAGTCTTTCCCGAGAGCGGCGCTCTACACAC 695
Qy 198 GlnGluLeuArgTyrSerValAlaLeuValLeuLysGluMetAla----- 212
Db 696 CAGGTGGCCACCAGGCTGTGGGCTGCATCATGGGGAGCTGTGAATGGTCCGCCCTT 755
Qy 213 -----AspLeuGluLysArgCys 218
Db 756 TTCCCGGGCAAGACGATATTGAACAGCTTTGC 788

RESULT 4

US-09-248-137-2
; Sequence 2, Application US/09248137
; Patent No. 6030788
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

LOCATION: 1...3049
US-09-221-017B-857
Alignment Scores:
Pred. No.: 2.84 Length: 3049
Score: 84.00 Matches: 53
Percent Similarity: 41.62% Conservativity: 29
Best Local Similarity: 26.90% Mismatches: 80
Query Match: 7.42% Indels: 35
DB: 4 Gaps: 10

US-09-904-568-4 (1-224) x US-09-221-017B-857 (1-3049)

QY 15 GlyAspSerHis-----SerGlyArgValSerPheLeu 25
DB 900 GGGGATTCGAGGGGATAGCTGCATATTGGGTACGGGCTCCAATTCCTGCTGTTCGAT 841
QY 26 GlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeuGlyAspLeuArg 45
DB 840 GGTCTGGGAGATCAAGCAACGTTTCCCACTGGGCTATATCTCGAGATGAAGGAGC 781
QY 46 SerThrPheArgLeuLeuLeuValSerSerLeuGlnGlyGluAsp----- 63
DB 780 GGAGCGGTATTGGGCGGCTG---TTCATCGGCGAGCTTCTCAAGGGGCGAGATGCCGAG 724
QY 64 -----CysArgAspGlyValGlnArgLeuGlyVal---SerAlaAsnLeuProGluGlu 80
DB 723 GGGCTGTGGAAGCTTCTCGAGGAGTATGGCTGACGTCTCCCATATATTCGAGAGT 664
QY 81 GlnLeu-----GlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGln 95
DB 663 GTCTATCGAAGCCTTTCCCAATCGCTTCTTGGCGGATTCCTCCCTTTATTCGGCGAG 604
QY 96 AlaLeuArgLeuProProThr-----SerLeuLeuProAspThrPheArgAspGlnLeu--- 113
DB 603 CATTTGGACATCCGCTGTATATAGCTTGTACAGAAATGTTTCGACGACTTCTTGTG 544
QY 114 -----GlnGluLeuLeuCysIleProGlnAspLeuValGlyAspLeuAla 127
DB 543 CGCAAGCTGCTCGATACATCTGCTGTATCTCCCTCCCTCCACTTCATCGGC----- 493
QY 128 SerValValPheGlySerGlnArgProLeuLeuAspSerValAlaGlnGlnGlyAla 147
DB 492 TCTGTAGCCTTT---CATTTACGGGAAGTGTTCAGTAGCGTATCAAAAAAGAGGCTTA 436
QY 148 TrpLeuProHisVal-AlaAspPheArgTrpArgValAspValAlaIleSerThrSeral 167
DB 435 ACCTTCGGATCGGTACTGCAAGTCCGATGGGAGGACTTATACAATATCACCAACAAT 376
QY 167 aLeuAlaArgSerLeu-----GlnProSerValLeuMetGlnLeuLys 181
DB 375 CATGTTTGAGAAGATTACCAACAGCCTTCCCTGTAGCAGGACTGGAA 327

RESULT 9
US-09-453-702B-194
; Sequence 194, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REFERENCE/DOCKET NUMBER: 27386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 9057
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 194:
US-09-453-702B-194
Alignment Scores:
Pred. No.: 15.5 Length: 9057
Score: 84.00 Matches: 53
Percent Similarity: 39.64% Conservativity: 35
Best Local Similarity: 23.87% Mismatches: 62
Query Match: 7.42% Indels: 72
DB: 4 Gaps: 12

US-09-904-568-4 (1-224) x US-09-453-702B-194 (1-9057)

QY 35 AlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeuLysPhe 54
DB 7014 GCATTTCCGCCCTGGTGTCTGTTGACGGCGCAGCAAGGCAAA----- 7061
QY 55 ValValSerSerLeuGlnGlyAspCysArgAspGlyValGlnArgLeuGlyValSer 74
DB 7062 ---GCACAACGGATACAAATCAGGAGTGTCTGTTTTCGCGCGTCCGCGATGTCGCC 7118
QY 75 AlaAsnLeuProGluGlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGln 94
DB 7119 GCACATATCCATGAA-----CTGGCAAAATCGCCAGCAGCTTAACCAAG 7160
QY 95 GlnAlaLeuArgLeuProProThrSerLeuLysProAspThrPhe----- 109
DB 7161 CAGGTCTACAGTCTATTGAACCGGACCTGAAACCGCGGTGATACCCCGCGGGCATGT 7220
QY 110 -----ArgAspGlnLeuGlnGlnCysIleProGlnAspLeuValGlyAsp----- 125
DB 7221 CCGCGCCGTAATGTCGTACCGGAGTTTTCGCTGCTCAAGATTCTTTTCATCCGCCAGS 7280
QY 126 -----LeuAlaSerValValPheGlySerGlnArgPro 136
DB 7281 GTCTTTATGGGACTATCACTATTATTGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 7340
QY 137 LeuLeu-----AspSerValAlaGlnGln-----GlyAlaTrp----- 148
DB 7341 GCGTTAAAGATCAGTTTTCATGCCCTTTTGTATGAGAAATACCGCGATATCTGGCTCAAT 7400
QY 149 -----LeuProHisValAlaAspPheArgTrpArgValAspValAlaIleSerThr--- 165
DB 7401 ATTCTGCTACCA-----CGGGTTTTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCT 7439
QY 166 SerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGln----- 179
DB 7440 GCAGCGCTTGCAACCGCA-----GGGGTAATATGACAGGGGCTTTTTCGCAACCCA 7490

```

Oy 180 -----Leuylsleuseraspjlyserla----- 187
Db 7491 ATGGCTGACCCAGGATTAATCTGGGGCTAGTACGGGCTTCGATTAATGCTGGCTTGA 7550
Oy 188 -----TyArgPheGluValaProThralaAlaSpheGlnGluLeuArgTyrSer 203
Db 7551 ATGTGCTGACCTTCTCTCCAGTGGTGTGCTGCTGATGAGCAAAATGATGTTGGC 7610
Oy 204 ValAla 205
Db 7611 ATTTGCC 7616

RESULT 10
US-08-996-083-2/c
; Sequence 2, 6124095
Patent No. 6124095
GENERAL INFORMATION:
APPLICANT: Magna, Holly
APPLICANT: Schaffer, Paul
APPLICANT: Lawton, Michael
APPLICANT: Yocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Murphy, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
FILE REFERENCE: PF-0420 US
CURRENT APPLICATION NUMBER: US/08/996, 083A
CURRENT FILING DATE: 1997-12-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 4183
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte Clone No. 6124095; 1388013
US-08-996-083-2

Alignment Scores:
Pred. No.: 6.2 Length: 4183
Score: 83.00 Matches: 49
Percent Similarity: 37.1% Conservative: 23
Best Local Similarity: 25.26% Mismatches: 79
Query Match: 7.33% Indels: 43
Gaps: 8
DB: 3

US-09-904-568-4 (1-224) x US-08-996-083-2 (1-4183)
Oy 5 GlyAlaAlaThrProTyrGluHisHisProGlyAspSerHisSerGlyValArgSer-Ph 24
Db 1078 GGGCCCCCTACCCTCACTGACCTGCGCTGTCCAGCAGGGTCCCATTTGTGAACACAGAGATT 1019
Oy 24 EleuGlyValaGlnLeuProProGlnValAlaAlaMetAlaArgLeuGlyAspLeuAs 44
Db 1018 TCTTGGGCGATGGGGGTCCCGAGGCTTTGGAGCAGAAAGTCACATCTTGGCCAGCCTTC 959
Oy 44 ParGSerThrPheArg-----LysLeuLeuLysPheValValSerSerle 59
Db 958 GCACCTCGGAGACTGACGGGTGTTTCCACGAGTACGGCTTCTCCAACTTATCATCAAGATGAT-- 901
Oy 59 uGlnGlyValAspArgValArgArgGlyValGlnArgLeuGlyValSerAlaAsnLeuProGln 79
Db 900 -----GGTACCAACAAGATGSA-----TCCGTTGGCGTGGGCTTGGCCCTCC-- 856
Oy 79 uGlnGlnLeuGlyValaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLe 99
Db 855 -----TTCAGTGAAGGATCCATCATCT--TGACAGGAAGCCATCCATCT 836
Oy 99 uProPro-----ThrSerLeuLysProAspThrThrPheArgArgGlnLeuGlnGluLe 116

```

```

Db      835 GGGCCCTATGTTGGCGGGCTGTCAGACAGACACCGAGC-----
Qy      116 ucys11eprcginaspylevaliglyaspheulalservalvalphneilysertinatgPr 136
Db      792 -----CCGAAAGGTTCCGG-----AGCATCGTGTGTCACAGTCCAG 752
Qy      136 oteuleuaspservalalagininglinglyalatrpleuProhisvalalaaappreAr 156
Db      751 GGTGTGTCGACGAGGAGAACCTGCGCTCCTAGCACTGTTGCCAGATGGGATGACCCG 692
Qy      156 gTParValaspValala-----lleserThrsAlaleuAlaarsertleuGl 173
Db      691 AGCCACGAGAGATGTTGTCGCCGCGATTACACAGTGTCAAGGCTGTCCACCTCGACACCGAG 632
Qy      173 nProserValleuMetGlnLeuLySleuSerApGlySer 186
Db      631 GCCGCACGCACTTCTGCGCTTCAGAGACGCGCCGGACA 592

RESULT 11
US-09-429-516-2/c
/ Sequence 2, Application US/09429516
/ Patent No. 6251389
/ GENERAL INFORMATION:
/ APPLICANT: Magna, Holly
/ APPLICANT: Schaffer, Paul
/ APPLICANT: Lawton, Michael
/ APPLICANT: Yocum, Sue
/ APPLICANT: Mitchell, Peter
/ APPLICANT: Hutchinson, Nancy
/ APPLICANT: Murry, Lynn E.
/ TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS
/ TITLE OF INVENTION: PHOXYDROLASE-2
/ NUMBER OF SEQUENCES: 3
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/429,516
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/996,083
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0420 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-855-0555
/ TELEFAX: 650-845-4166
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4183 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: SATPF1008
/ CLONE: 1388013
/ US-09-429-516-2

Alignment Scores: 6.2 Length: 4183
red. No.:

```

```
Score: 83.00 Matches: 49
Percent Similarity: 37.11% Conservative: 23
Best Local Similarity: 25.26% Mismatches: 79
Query Match: 7.33% Indels: 43
DB: 4 Gaps: 8

US-09-904-568-4 (1-224) x US-09-429-516-2 (1-4183)

Qy 5 GlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGlyArgValSer-Ph 24
Db 1078 GGGCCCCGTACCATGAGCTCCCTGTGCAGCAGGTCCTTCTCCAACTTATCAAGGATGAT 1019

Qy 24 eLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAs 44
Db 1018 TCTTGGGCATGGGTCCCGGAGGCTTTCACACAGAAAGTCACATTCGGCCAGCCTCTC 959

Qy 44 pArgSerThrPheArg-----LysLeuLeuLysPheValValSerSerLe 59
Db 958 GCACCTCGGACTCAGGGTGTTCACACAGGTACGGTCTCTCCAACTTATCAAGGATGAT-- 901

Qy 59 uGlnGlyGluAspCysAtqAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProG 79
Db 900 -----GGTGACCACAGAGATGGA-----TCCGTGTGCCCTGGCGCTGCCCTCCC-- 856

Qy 79 uGlnGlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLe 99
Db 855 -----TCGACAGAGCCATCCATCT 836

Qy 99 uProPro-----ThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLute 116
Db 835 GGGCCCTGTATGTGGCGCGCTGTGCAGCAGACACACAGGCAC----- 793

Qy 116 uCysIleProGlnAspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgPr 136
Db 792 -----CCGGAAGTTCCGTG-----AGCATCGTGTGGTCCACAGTCCCGAG 752

Qy 136 oLeuLeuAspSerValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheAr 156
Db 751 GCTGGTCTCCAGGGAGACCTCGCTCCTAGCAGTGGTGGTCCAGATGGGTGACCAACG 692

Qy 156 gTrpArgValAspValala-----IleSerThrSerAlaLeuAlaArgSerLeuGl 173
Db 691 AGCCAGGAGGATGTGTCTCGCGCATTCACAGGTGTCAAGGTGCACCCCTGGACACCGAG 632

Qy 173 nProSerValLeuMetClnLeuLysLeuSerAspGlySer 186
Db 631 GCCGACGCACTTCTGCGCTCCAGAGGACGCCCGGACA 592

RESULT 12
US-08-241-766-12/c
; Sequence 12, Application US/08241766
; Patent No. 568590
; GENERAL INFORMATION:
; APPLICANT: JACOBS, W. R.
; APPLICANT: COLLINS, D. M.
; APPLICANT: BANERJEE, A.
; APPLICANT: DELISLE, G. W.
; APPLICANT: WILSON, T. M.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/241,766
; FILING DATE: 12-MAY-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MONROY, GLADYS H.
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 25237-20003.20
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 494..1234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1256..2062
; US-08-241-766-12

Alignment Scores:
Pred. No.: 3.1 Length: 2232
Score: 82.00 Matches: 51
Percent Similarity: 36.60% Conservative: 20
Best Local Similarity: 26.29% Mismatches: 59
Query Match: 7.24% Indels: 64
DB: 1 Gaps: 10

US-09-904-568-4 (1-224) x US-08-241-766-12 (1-2232)

Qy 13 HisProGlyAspSerHisSer---GlyArgValSerPheLeuGlyAlaGlnLeuProPro 31
Db 1564 CATCCCGGTCTCGGCATGAACCAATCGCATGCACAC-----CCCGTC 1520

Qy 32 GluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeu 51
Db 1519 GAGCTTGTTCGCCCGCCGATCGCTCGGTTCAGCAGCGGGCGCTTCCCG--CAGCCGTC 1460

Qy 52 LeuLysPheValValSerSerLeuGlnGlyGluAspCysArgAspGlyValGlnArgLeu 71
Db 1459 CTCGTTTTCAGCTCGAGTTCGAGCAGCGGGCGCTTCCCG--CAGCCGTC 1409

Qy 72 GlyValSerAlaAsnLeuPro-----GluGlnLeuGlyAlaLeu 85
Db 1408 GGTGATGCGCTGAATCAGCCGCGAGCGCGTTCGAACCCCGTGAGCACAGCTGGCGCCCTG 1349

Qy 86 LeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSerLeuLys 105
Db 1348 CTCCTGGGTACCCGTGGATGTGAACCGCATGCACAGTCGCGTATGATTCGCTAAC 1289

Qy 106 ProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln----- 120
Db 1288 CAGAATCCGTTT-----GCCGTCCAGCAGCTCTGTCTGTCGCTCTGTGTGTGTC 1235

Qy 121 -----AspLeuValGlyAspLeuAlaSerVal 129
Db 1234 GTGGCCCATACCCATCCCGCTCGACCGGGATGACCGCAGGAGATATAGCTCGCAT 1175

Qy 130 ValPheGlySer----- 133
Db 1174 CTC---GGAAGCCAGGAAGCTACACCCCGGCGGCTCGCGGGGGTCCGACCCGCTT 1118

Qy 134 -----GlnArgProLeuLeuAspSerValAlaGlnGlnGlnGlyAlaTrp 148
Db 1117 CGCTGGGATAAATTGACGCGCCCTGCTGAATCCGCTCATCCAGCGCGCGGT----- 1064

Qy 149 LeuProHisValAlaAspPheArgTrpArgValAspValala 162
```

Db 1063 -----CATATCGGT-----GTGATGTAGCC 1043
RESULT 13
US-08-491-146-1/c
Sequence 1, Application US/08491146
Patent No. 5556778
GENERAL INFORMATION:
APPLICANT: Sacchetti et al
TITLE OF INVENTION: INHA CRYSTALS AND THREE
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Amerer, Rochstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/491,146
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/307,376
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: George, Kenneth P
REGISTRATION NUMBER: 30,259
REFERENCE/DOCKET NUMBER: 96700/296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3120
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: DNA
DESCRIPTION:
HYPOTHEICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: inhA operon
ORGANISM: M tuberculosis
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: M tuberculosis
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: No. 5556778
AUTHORS:
TITLE:

JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-491-146-1
Alignment Scores:
Pred. No.: 5.23 Length: 3120
Score: 82.00 Matches: 51
Percent Similarity: 36.60% Conservative: 20
Best Local Similarity: 26.29% Mismatches: 59
Query Match: 7.24% Indels: 64
DB: 1 Gaps: 10
US-09-904-568-4 (1-224) x US-08-491-146-1 (1-3120)
QY 13 HisProGlyAspSerHisSer---GlyArgValSerPheLeuGlyAlaGlnLeuProPro 31
Db 1294 CATCCGCTCGCGCATGAACCAATGATGACACAC-----CCGCTC 1250
QY 32 GluValAlaAlaMetValArgLeuGlyAspLeuSpArgSerThrPheArgLysLeu 51
Db 1249 GAGCTTGTGGCCCGCCCGATCGCTCGTCACCGCGCCGCAAGCTGGCCAGTGTCTC 1190
QY 52 LeuLysPheValValSerSerLeuGlnGlyLysArgGlyValGlnArgLeu 71
Db 1189 CTCCTTTGCACGTCGAGTTGAGACAGCGGGCCCTTGTGG-----CACCCGCTC 1139
QY 72 GlyValSerAlaAsnLeuPro-----GluGlnGlnLeuGlyAlaLeu 85
Db 1138 GGTATGCGCTGATTCAGCCGACCGGTGGAACCGGTGACACCGCTGGCGCTG 1079
QY 86 LeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSerLeuLys 105
Db 1078 CTCCTGGCTACCCGCTCGATGTGAACGGCATCGACGATCGGTGATTCGCTAAC 1019
QY 106 ProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln----- 120
Db 1018 CAGATCCGTT-----GCCGTCCAGACAGTCTGTATGTGCTTGTGTGTCA 965
QY 121 -----AspLeuValGlyAspLeuAlaSerVal 129
Db 964 GTGGCCCATACCATGCGCGCGTGCACCGGGATGACCGACCGAGATATAGCTCGCATC 905
QY 130 ValPheGlySer----- 133
Db 904 CTC---GAAAGCCAGGAAGTGAACCCCGGAGCTCGCGGGGTGCCGACCCGCTT 848
QY 134 -----GlnArgProLeuLeuAspSerValAlaGlnGlnGlnGlyAlaTrp 148
Db 847 CGCTGGATTAATTGACGCGCCCTCTGTGATTCGCTCATCCAGCGCGGGGT----- 794
QY 149 LeuProHisValAlaAspPheArgTrpArgValAlaValAla 162
Db 793 -----CATATCGGT-----GTGATGTAGCC 773
RESULT 14
US-08-241-766-1/c
Sequence 11, Application US/08241766
Patent No. 5686590
GENERAL INFORMATION:
APPLICANT: JACOBS, W. R.
APPLICANT: COLLINS, D. M.
APPLICANT: BANERJEE, A. M.
APPLICANT: DELISLE, G. W.
APPLICANT: WILSON, T. M.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
AND TREATING MYCOBACTERIAL INFECTIONS USING AN INHA AGENT
NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORRISON & FOERSTER
;; STREET: 755 Page Mill Road
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304-1018

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/241,766
;; FILING DATE: 12-MAY-1994

;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: MONROY, GLADYS H.
;; REGISTRATION NUMBER: 32,430
;; REFERENCE/DOCKET NUMBER: 25237-20003.20
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 813-5600
;; TELEFAX: (415) 494-0792
;; TELEX: 706141

;; INFORMATION FOR SEQ ID NO: 11:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3120 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-241-766-11

Alignment Scores: 5.23 Length: 3120
Pred. No.: 82.00 Matches: 51
Score: 36.60% Conservative: 20
Percent Similarity: 26.29% Mismatches: 59
Best Local Similarity: 7.24% Indels: 64
Query Match: 1 Gaps: 10
DB:

US-09-904-568-4 (1-224) x US-08-241-766-11 (1-3120)

QY 13 HisProGlyAspSerHisSer---GlyArgValSerPheLeuGlyAlaGlnLeuProPro 31
Db 1294 CATCCCGCTCGGCATGATACCACTGATGCACAC-----CCGCTC 1250
QY 32 GluValAlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeu 51
Db 1249 GAGCTTGTTCGCCGCCCGATCGCTCGGTCAACCGCGCGGCAAGCTGCCAGGTGCTC 1190
QY 52 LeuLysPheValValSerSerLeuGlnGlyGluAspCysArgAspGlyValGlnArgLeu 71
Db 1189 CTCGTTTTCACGTCGAGTTCGAGCAGCGGGGCTTTGCCGG-----CAGCCGGTC 1139
QY 72 GlyValSerAlaLeuLeuPro-----GluGluGlnLeuGlyAlaLeu 85
Db 1138 GGTGATCGCGTGTGATACCGCGCCCGGTGCAACCGGTGAGCAGCAGCTGGGGCGCTG 1079
QY 86 LeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuProThrSerLeuLys 105
Db 1078 CTCCTGGGTACCGCTGCGATGTAACCGGATCGACGAGTGGTGATTCGCTAAC 1019
QY 106 ProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln----- 120
Db 1018 CAGAAATCCGTTT-----GCCGTCCAGCAGTCTCTGTCATGTGCGCTCTGTGTGTGTC 965
QY 121 -----AspLeuValGlyAspLeuAlaSerVal 129
Db 964 GTGGCCCATACCCATGCCCGCTCGACCGGGATCACCGCAGGAGATAGCTCGCATC 905
QY 130 ValPheGlySer----- 133
Db 904 CTC---GGAAGCCAGGAAGCTGACACCCCGCGGACCTCGCGGGGGTTCGACCCGCTT 848

QY 134 -----GlnArgProLeuLeuAspSerValAlaGlnGlnGlnGlyAlaTrp 148
Db 847 CGCTGGGATAAATTGCAGCGCCCTCTGTAATCCGCTCATCCAGCGCGGGT----- 794
QY 149 LeuProHisValAlaAspPheArgTrpArgValAspValAla 162
Db 793 -----CATATCCGT-----GTCGATGTAGCC 773

RESULT 15
US-08-234-011-1/c
; Sequence 1, Application US/08234011
; Patent No. 5702935
; GENERAL INFORMATION:
; APPLICANT: Sacchettini et al
; TITLE OF INVENTION: METHOD AND COMPOUNDS FOR INHIBITING
; TITLE OF INVENTION: LIPID BIOSYNTHESIS OF BACTERIA AND
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/234,011
; FILING DATE: No. 5702935 yet assigned
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: none
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: 96700/271
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 3120
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; MOLECULE TYPE: DNA
; DESCRIPTION: No
; HYPOTHETICAL: No
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: inhA operon
; ORGANISM: M tuberculosis
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLSE:
; IMMEDIATE SOURCE: M tuberculosis
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:

IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: NO. 5702935e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-234-011-1

Alignment Scores:
Pred. No.: 5.23 Length: 3120
Score: 82.00 Matches: 51
Percent Similarity: 36.60% Conservative: 20
Best Local Similarity: 26.22% Mismatches: 59
Query Match: 7.24% Indels: 64
DB: 1 Gaps: 10

US-09-904-568-4 (1:224) x US-08-234-011-1 (1-3120)

QY 13 HisProG1AspSerHisSer---GlyArgValSerPheLeuGlyAlaGlnLeuProPro 31
DB 1294 CATCCGGGCTCGCGGATCAACCCCAATCGAATGCACAC-----CCCGTC 1250

QY 32 GluValAlaAlaMecAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeu 51
DB 1249 GAGCTTGTTGCCCGCCCGCATCGCTCGGTACCCGCCCGCAAGCTGGCCAGGTCTC 1190

QY 52 LeuLysPheValValSerSerLeuGlnGlyLysArgAspGlyValGlnArgLeu 71
DB 1189 CTCGTTTTCACGTCGAGTTCGAGCAGCGGGCCTTGGCCG-----CAGCCGGTC 1139

QY 72 GlyValSerAlaAsnLeuPro-----GluGlnGlnLeuGlyAlaLeu 85
DB 1138 GGTGATGCCCTGATCAACGCGCAGCCCGGTGACCCGGTGAACAGCTGGCGCCCTG 1079

QY 86 LeuAlaGlyMecHisThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSerLeuLys 105
DB 1078 CTCCTGGGCTACCGCGGATGTGAACGCCATCGACGAGTGGATGATTCGCTAAC 1019

QY 106 ProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln----- 120
DB 1018 CAGAAATCCGTT-----GCCGTCAGACATCTGTGATGTCCTGTGTGTGTCA 965

QY 121 -----AspLeuValGlyAspLeuAlaSerVal 129
DB 964 GTGGCCCATACCATCGCGCCCTGCACCGGGATGACCGCAGAGATATAGCTGCATC 905

QY 130 ValPheGlySer----- 133
DB 904 CTC---GGAAGCCAGAGACTGACACCCCGCGACCTCGGGGGGTGCCACCCGCTT 848

QY 134 -----GlyArgProLeuLeuAspSerValAlaGlnGlnGlnGlyAlaTrp 148
DB 847 CGCTGGGATTAATTGCACGCGCCCTGCTGATCCGCTCATTCACGCGCGGGT----- 794

QY 149 LeuProHisValAlaAspPheArgTrpArgValAspValAla 162
DB 793 -----CATATCGGT-----GTGATGTAGCC 773

Search completed: December 7, 2002, 04:28:41
Job time: 63 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 7, 2002, 03:13:40 ; Search time 36.5 Seconds
(without alignments)
2393.012 Million cell updates/sec

Title: US-09-904-568-4
Perfect score: 1132
Sequence: 1 MSAGVGAATPYLHPGDSHSG.....ALVLKEMADLEKRCERRIQD 224

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frames_p2n.model -DEV=xlpl
-Q=/cg2_1/USPTO.spool/US09904568/runat_25112002_143307_20020/app_query.fasta_1.782
-DB=Published Applications NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40 cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09904568 -CGN 1 1 36 -runat_25112002_143307_20020
-NCPUS=6 -ICPU=3 -NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1132	100.0	1355	10	US-09-904-568-3 Sequence 3, Appli
2	1084	95.8	1258	10	US-09-925-300-474 Sequence 474, App
3	906	80.0	1100	10	US-09-904-568-1 Sequence 1, Appli
4	461	40.7	280	10	US-09-864-761-19433 Sequence 19433, A

5	355.5	31.4	386	10	US-09-783-590-8595	Sequence 8595, Ap
6	314	27.7	406	10	US-09-880-107-1264	Sequence 1264, Ap
7	289	25.5	401	10	US-09-864-761-2717	Sequence 2717, Ap
8	219	19.3	212	10	US-09-783-590-8782	Sequence 8782, Ap
9	188.5	16.7	192	10	US-09-783-590-8586	Sequence 8586, Ap
10	168	14.8	188	10	US-09-783-590-8742	Sequence 8742, Ap
11	97	8.6	945	10	US-09-815-242-7826	Sequence 7826, Ap
12	93.5	8.3	751	10	US-09-848-915-2	Sequence 2, Appli
13	93.5	8.3	751	12	US-10-044-090-766	Sequence 1762, App
14	89.5	7.9	1458	10	US-09-925-301-172	Sequence 172, App
15	89.5	7.9	1749	10	US-09-870-162A-22	Sequence 22, Appl
16	89.5	7.9	14272	10	US-09-870-162A-23	Sequence 23, Appl
17	86	7.6	32185	10	US-09-764-877-3171	Sequence 3171, Ap
18	85.5	7.6	1491	10	US-09-796-393-1	Sequence 1, Appli
19	83.5	7.4	433	10	US-09-983-965-3790	Sequence 3790, Ap
20	83.5	7.4	830	10	US-09-925-301-301	Sequence 301, App
21	83	7.3	4183	10	US-09-757-716-2	Sequence 2, Appli
22	82	7.2	460	10	US-09-983-965-1873	Sequence 1873, Ap
23	82	7.2	3497	10	US-09-761-569-5	Sequence 5, Appli
24	82	7.2	3553	10	US-09-761-569-7	Sequence 7, Appli
25	82	7.2	3576	10	US-09-761-569-9	Sequence 9, Appli
26	81.5	7.2	1060	10	US-09-768-826-14	Sequence 14, Appl
27	81.5	7.2	5793	10	US-09-880-107-2109	Sequence 2109, Ap
28	81	7.2	1997	10	US-09-925-301-555	Sequence 555, App
29	80	7.1	482	10	US-09-738-973-481	Sequence 481, App
30	80	7.1	1002	10	US-09-815-242-7774	Sequence 7774, Ap
31	80	7.1	4117	10	US-09-823-356-26	Sequence 26, Appl
32	79	7.0	1584	9	US-09-712-363-10	Sequence 10, Appl
33	79	7.0	2187	10	US-09-508-8058-7	Sequence 7, Appli
34	79	7.0	48667	10	US-09-822-268A-3	Sequence 3, Appli
35	78	6.9	2364	12	US-10-044-090-37	Sequence 37, Appl
36	77.5	6.8	14486	9	US-09-764-868-1504	Sequence 1504, Ap
37	77.5	6.8	80959	9	US-09-858-546-3	Sequence 3, Appli
38	77	6.8	368	10	US-09-738-973-570	Sequence 570, App
39	77	6.8	399	10	US-09-880-107-1037	Sequence 1037, Ap
40	77	6.8	847	10	US-09-960-253-183	Sequence 183, App
41	77	6.8	1401	10	US-09-825-414-19	Sequence 19, Appl
42	77	6.8	2067	10	US-09-728-952-37	Sequence 37, Appl
43	77	6.8	11458	10	US-09-825-414-18	Sequence 18, Appl
44	76.5	6.8	586	10	US-09-822-849A-525	Sequence 525, App
45	76.5	6.8	2464	10	US-09-848-035-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-904-568-3
; Sequence 3, Application US/09904568
; Patent No. US20020115604A1
; GENERAL INFORMATION:
; APPLICANT: TREMBLAY, Johanne
; APPLICANT: HAMET, Pavel
; APPLICANT: LEWANCZUK, Richard
; APPLICANT: GOSSARD, Francis
; TITLE OF INVENTION: A No. US20020115604A1
; FILE OF INVENTION: (HcArg)
; FILE REFERENCE: 12725.9
; CURRENT APPLICATION NUMBER: US/09/904,568
; CURRENT FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (295)..(966)
US-09-904-568-3

Alignment Scores: 3.06e-125 Length: 1355
Pred. No.: 1132.00 Matches: 224
Score:

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Matchn: 100.00% Indels: 0
Ds: 10 Gaps: 0
```

US-09-904-568-4 (1-1224) X US-09-904-568-3 (1-1355)

QY	1	MeSeSerAlaValAlaValAlaThrProCysLeuHisAspProGlyAspSerHisSerGly	20
Db	295	ATGTCGTGCTGGGGGCTGCMAACTCCAACTGCATCTCTGGTGATGTCACAGTGGC	354
QY	21	ArgValSerPheLeuGlyAlaGlnLeuProProGluValAlaAlaMetValArgLeuLeu	40
Db	355	CGAGGACGTTCTTGGGGGCCAGCTTCCTCCAGAGGGGCGACGCAATGGCCCGGCTCTA	414
QY	41	GlyAspLeuAspArgSerThrPheArgValLeuLeuLysPheValValSerSerLeuGln	60
Db	415	GGGAGCCTAGACAGGAGGACGTTCCAGAAAGTTGCTGAAGTTTGAGTCCAGAGCTTCAG	474
QY	61	GlyIleAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGln	80
Db	475	GGGGAGCATCTCCGAGAGCGGTGTGACGCTCTTGGGGTCAAGCCCAACTGCGGAGAG	534
QY	81	GlnLeuGlyValAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro	100
Db	535	CAGCTGGGTGCTTCGTGTGGCAGGATGCACACTGCTCCAGCAGGGCTTCGCTGCC	594
QY	101	ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGlnLeuCysIleProGln	120
Db	595	CCCAACAGCTCGAAGCCCTGCACCTTCAGGGACCAAGCTCCAGAGAGCTCTGCATCCCA	654
QY	121	AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer	140
Db	655	GACCTGGTGGGGACTTGTGGCCAGGGTGGTAATTGGAGGCCAGCGGCCCTCTTGATTCT	714
QY	141	ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp	160
Db	715	GTGGGCCAGCAGCAGGGGGCTGTGGCTCCGATCTTCTCACTTTCGGTGGGGGTGAT	774
QY	161	ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu	180
Db	775	GTACCAACTCTCCACAGGCGCTTCGTGCTGCCCTCTGCAGCGGAGCGTCTGATGCAAGCTG	834
QY	181	LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGlnLeu	200
Db	835	AAGCTTTCAGTGGGTCAAGCATACCGCTTTGAGGTCCCCACAGCCAAATTCCAGAGACTG	894
QY	201	ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGlnLysArgCysGluArg	220
Db	895	CGGTACACCGCGGCGCTGTGTCCTAAAGAGATGGCAATCTGGAGAAAGAGGTGAGCGC	954
QY	221	ArgLeuGlnAsp	224
Db	955	AGACTGCAGGAC	966

RESULT 2
US-09-925-300-474
Sequence 474, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruden,
FILE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925.300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 474

```

? LENGTH: 1258
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (36)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (528)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (726)
? OTHER INFORMATION: n equals a,t,g, or c
? OS-925-300-474

```

Alignment Scores:

Pred. No.:	1,53e-120	Length:	125e
Score:	1084.00	Matches:	215
Percent Similarity:	96.88	Conservative:	2
Best Local Similarity:	95.98	Mismatches:	7
Query Match:	95.76	Indels:	0
DB:	10	Gaps:	0

US-09-904-568-4 (1-224) X US-09-925-300-474 (1-1258)

QY	1	MetSerAlaValGlyValAlaIleThrProTyrLeuIleHisAspProGlyAspSerHisLeuSerGly	20
Db	330	ATGTCGTGTCGTGGGRCGCGCAACTCCATACCTGATCATCTGTGTATGTCACAGATGCC	388
QY	21	ArgValSerPheLeuGlyValGlnLeuProProGlyValAlaAlaMetAlaArgLeuLeu	40
Db	390	CGAATGAGTTTCCTTGGGGGGCCACACTTCCTCCAGAGGTGGCAGCAATGGCCCGCTACTA	448
QY	41	GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValAspSerLeuGln	60
Db	450	GGGGAACCTAGACAGACAGACGTTGAGAAAGTTGCTGAAGTTGTGTGCAGAGCCTCAG	508
QY	61	GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaMetLeuProGluGln	80
Db	510	GGGAGAGACTGCCAGACAGANTGCTGCACAGCTCTGGGGTCAAGCCGCAACTGGCGAGAG	568
QY	81	GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro	100
Db	570	CAGCTGGGTGTCCTGCTGGGAGGCGATGCACACACTGCTCCAGACAGGCCCTCCGCTGCC	628
QY	101	ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLysLeuCysIleProGln	120
Db	630	CCCAACGACGCTGAAGCCTGACACCTTCAGGGACCAAGTCCAGAGACTCTGCAATCCCCAA	688
QY	121	AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer	140
Db	690	GACCTGTGCGGGACTTGGCCAGCGTGCTGTTTGGGAGACCAAGCGGCCCTCCCTTGATTCT	748
QY	141	ValAlaGlnGlnGlnGlnValATrPheLeuProHisValAlaAspPheArgTrpArgValAsp	160
Db	750	GTCGCCACAGACAGAGGGGGCGCTGCTGCCCACTTGTCTACTCTTGCTGGTGGCGGTGAT	808
QY	161	ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu	180
Db	810	GTAGCAATCTCCACAGTGGCCCTGGCTCGCTCCCTGGAGCGGCGTCTGATGACGCTG	868
QY	181	LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGlnLeu	200
Db	870	AAGCTTTCACATGGGTGTCAGATACCGCTTGTAGAGTCCCAACCAAGTTCACGAGCTG	928
QY	201	ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuLysArgCysGluArg	220
Db	930	CGGTACAGCGTGGCCCTGGTCTTAAAGAGATGCGAGATCTGGAAGAAGGTGTACGCC	988
QY	221	ArgLeuGlnAsp	224
Db	990	AGACTGCAGAC	1001

RESULT 3

US-09-904-568-1
; Sequence 1, Application US/09904568
; Patent No. US20020115604A1
; GENERAL INFORMATION:
; APPLICANT: TREMBLAY, Jeanne
; APPLICANT: HAMEZ, Pavel
; APPLICANT: LEWANCZUK, Richard
; APPLICANT: GOSSARD, Francis
; TITLE OF INVENTION: A No. US20020115604A1el Hypertension Related Calcium Regulated Ge
; TITLE OF INVENTION: (HcARG)
; FILE REFERENCE: 12725.9
; CURRENT APPLICATION NUMBER: US/09/904.568
; CURRENT FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (132)..(803)
US-09-904-568-1

Alignment Scores:
Pred. No.: 2,54e-99 Length: 1100
Score: 906.00 Matches: 178
Percent Similarity: 89.73% Conservative: 23
Best Local Similarity: 79.46% Mismatches: 23
Query Match: 80.04% Indels: 0
DB: 10 Gaps: 0

US-09-904-568-4 (1-224) x US-09-904-568-1 (1-1100)

Qy	1	MetSerAlaValGlyAlaAlaThrProTyrLeuHisHieProGlyAspSerHisSerGly	20
Db	132	ATGCTGCTTTGGGGCTGCAGCTCCATCTTGCACCATCCCGCTGACAGTCACAGTGGC	191
Qy	21	ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu	40
Db	192	CGGGTCAGTTTCCTGGGTTCCTCCAGAGCTTCCAGAGTGCAGCGCTGCTCAGTCTTG	251
Qy	41	GlyAspLeuAspArgSerThrPheArgLysLeuLysPheValValSerSerLeuGln	60
Db	252	AAGACTTAGCAGGAGCACCCTTCAGAAAGTTGTTGAACTGTAGTCGGGGCCCTGCAT	311
Qy	61	GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu	80
Db	312	GGCAAGAGCTGCAGAGAAGCTGTGGAGCACTTGGTCCAGCGCCCAACCTGTGAGAAGAG	371
Qy	81	GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro	100
Db	372	CGTTCGCCCTGCTGCTGGCGGCGCACACACACCTGCTCCAGCAGGCTCTCCGGCTGCC	431
Qy	101	ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln	120
Db	432	CTGTAGTCTAAGCCAGATGCTTCCAGAGAGCTCCAGAACTTGCATTCCTCAG	491
Qy	121	AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer	140
Db	492	GATCTAATGGAGATTGGCCAGTTTGGCATTTGGGAGTCAACGCCCTCTCTCGACTCT	551
Qy	141	ValAlaGlnGlnGlnGlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp	160
Db	552	GTAGCCCAACAGCAGGATCTCGCTCCCTCAGTGTCTTACTTCCGGTGGCGGTGGAC	611
Qy	161	ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu	180
Db	612	GTGCCCATCTCAACCGAGCTCAGTCCCGTCCCTCGAACCGAGTGTCTCATGCAGCTG	671
Qy	181	LysLeuSerAspGlySerAlaTrpArgPheGluValProThrAlaLysPheGlnGluLeu	200

Db	672	AAGCTCACAGATGGATCTGCACACCGCTTCGAGGTGCCCATAGCCAAATTTTCAGAGCTG	731
Qy	201	ArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluArg	220
Db	732	CGTACAGTGTAGCTTGGTCTTAAAGAGATGGCAGAACTGGAGAGAAAGTGTGAGCGC	791
Qy	221	ArgLeuGlnAsp	224
Db	792	AAACTGCAGGAC	803

RESULT 4

US-09-864-761-19433
; Sequence 19433, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19433
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF146367.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 9.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3

US-09-904-568-4 (1-224) x US-09-783-590-8595 (1-386)

QY 26 GlyAlaGlnLeuProGluValAlaAlaMetAlaAArgLeuLeuGlyAspLeuAspArg 45
DB 13 GCGTGNCAAGATTGGGCAAGTGGCAGCAATGGCCCGCTACTAGGGACCTAGACATG 72
QY 46 SerThrPheArgLysLeuLeuLysPheValSerSerLeuGlnGlyGluAspCysArg 65
DB 73 ANCACGTTCCAGAAAGTTCTGAAGTTTGTGGTCAGCAGCCTGCAGGGGGAGGACTGCCGA 132
QY 66 AspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlnLeuGlyAlaLeu 85
DB 133 AAGNTGTGCAGCGCTTGGGGTCCANCCCAACCTGNCAGGAGCAGCAGCTGGGTNCCCTG 192
QY 86 LeuAlaGlyMetHisThrLeuLeuGlnAlaLeuArgLeuProThrSerLeuLys 105
DB 193 CTGGCAGGATGCACACTGCTCCANCAAGNCCTCCGTNNCCCCCACCAGCCTGAAG 252
QY 106 ProAspThrPheArgAsp-GlnLeuGlnGluLeuCysIleProGlnAsp-LeuValGly 125
DB 253 CCTGAAACCTTCAGGGAACCAAGTCCAGGAGCTCTGAATCCCCCAAGAGCCTGGTCGGG 312
QY 125 sPleuAlaSerValPhe---GlySerGlnArgProLeuLeuAspSerValAlaGlnG 144
DB 313 AACTTGNCAAGTGTATTGGGAAGCAGGCGCCCTCGTTGATTCTTTGGGCCANN 372
QY 144 InGlnGlyAlaTrp 148
DB 373 AGGAGGGGGCTGG 386

RESULT 6

US-09-880-107-1264/c
; Sequence 1264, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1264
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA464962

Alignment Scores:
Pred. No.: 4,35e-29 Length: 406
Score: 314.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.74% Indels: 0
DB: 10 Gaps: 0

US-09-904-568-4 (1-224) x US-09-880-107-1264 (1-406)

QY 160 AspValAlaIleSerThrSerAlaLeuAlaAArgSerLeuGlnProSerValLeuMetGln 179
DB 406 GATGTAGCAATCTCCACCAAGTGCCTGGTCTCCTGCAGCCGAGCGCTCTGATGCAG 347
QY 180 LeuLysLeuSerAspGlySerAlaTyArgPheGluValProThrAlaLysPheGlnGlu 199

DB 346 CTGAAGCTTTTCAGATGGTTCAGCATACCGCTTTGAGGTCCCCACAGCAAGTTCCAGGAG 287
QY 200 LeuArgTyrSerValAlaLeuValLeuLysGluMetAlaAAspLeuGluLysArgCysGlu 219
DB 286 CTGCGGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAG 227
QY 220 ArgArgLeuGlnAsp 224
DB 226 CCGACACTGCAGGAC 212

RESULT 7

US-09-864-761-2717
; Sequence 2717, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006659
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2717
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF146367.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 9.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
US-09-864-761-2717

Alignment Scores:
Pred. No.: 4,2e-26 Length: 401
Score: 289.00 Matches: 57
Percent Similarity: 98.28% Conservative: 0
Best Local Similarity: 98.28% Mismatches: 1
Query Match: 25.35% Indels: 0
DB: Gaps: 0

US-09-904-568-4 (1-224) x US-09-864-761-2717 (1-401)

OY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisProGlyAspSerHisSerGly 20
DB 227 ATGTCTGCTGGGGAGCTGCACTCCATACCTGCATCATCTGGTATAGTCACAGTGGC 286
OY 21 ArgValSerPheLeuGlyAlaGlnLeuProProGluValAlaAlaMetAlaArgLeu 40
DB 287 CGAGTGAAGTTTCTGGGGCCCAAGCTTCTCCAGAGTGGCAGCAATGCCCCGCTACTA 346
OY 41 GlyAspLeuAspArgSerThrPheArgIysLeuLeuLysPheValIserSer 58
DB 347 GGGGACCTAGACAGAGCAGCTTCAGAAAGTTGCTGAAGTTGTGGTCAGCAGC 400

RESULT 8

US-09-783-590-8782
Sequence 8782, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:

APPLICANT: Dillion, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8782
LENGTH: 212
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (63)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (190)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (195)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (198)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (200)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (201)
OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc feature
LOCATION: (206)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (207)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-8782

Alignment Scores:
Pred. No.: 4,01e-18 Length: 212
Score: 219.00 Matches: 55
Percent Similarity: 80.88% Conservative: 0
Best Local Similarity: 80.88% Mismatches: 10
Query Match: 19.35% Indels: 3
DB: Gaps: 0

US-09-904-568-4 (1-224) x US-09-783-590-8782 (1-212)

OY 33 ValAlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgIysLeu 52
DB 7 GTGGCAGCAATGGCCCGGCTACTAGGGGACCTAGACATGAGCACGCTTCAGAAAGTTNCTG 66
OY 53 LysPheValIserSerLeuGlnGlyIubspCysArgAspGlyValGlnArgLeuGly 72
DB 67 AAGTTTGTGTCAGCAGCTTCAGGGGAGGAGGAGCAGCTGGGTCCTCTGGCAGGATGACACA 126
OY 73 ValSerAlaAsnLeuProGluGluGlnLeuGlyAlaLeuLeuAlaGlyMetHisThr 91
DB 127 GTCAGCCGCCCAACCTGGCGGAGAGCCAGCTGGGTCCTCTGGCAGGATGACACA 186
OY 92 LeuLeuGlnGlnAlaLeu 97
DB 187 TTGNTCCANGANTNNCTC 204

RESULT 9

US-09-783-590-8586
Sequence 8586, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:

APPLICANT: Dillion, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8586
LENGTH: 192
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (3)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (12)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (55)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (144)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (151)

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (153)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (174)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (178)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (179)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (183)
OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-8586

Alignment Scores:

Pred. No.:	1,55e-14	Length:	192
Score:	188.50	Matches:	49
Percent Similarity:	77.78%	Conservative:	0
Best Local Similarity:	77.78%	Mismatches:	11
Query Match:	16.65%	Indels:	3
DB:	10	Gaps:	1

US-09-904-568-4 (1-224) x US-09-783-590-8586 (1-192)

QY 33 ValAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeuLeu 52
|||
Db 8 GTGGNAGCAATGCCCGGCTACTAGGGACCTAGACATGACGACGTTNAGAAAGTTGCTG 67
QY 53 LysPheValValSerLeuGlnGlyGluAspCysArgAspGlyValGlnArgLeuGly 72
|||
Db 68 AAGTTTGTGTCAGACGCTGCGAGGGAGGAGTGGCGAAAGGTGTGCCAGCGCTTTGGG 127
QY 73 ValSerAla-AsnLeuProGlu-GluGlnLeuGlyAlaLeuLeuAla---GlyMetHist 91
|||
Db 128 GTGAGCGCCCAACTGTCGCGAGGNGNCAGCTGGTGGCTGCTGGANAGNNATGNCACA 187
QY 91 hrLeu 92
|||
Db 188 CACTG 192

RESULT 10

US-09-783-590-8742
Sequence 8742, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Haseltine, William A.
APPLICANT: Dillon, Patrick J.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16, 2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8742
LENGTH: 188
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (3)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature

LOCATION: (41)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (47)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (49)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (56)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (101)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (145)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (151)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (152)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (154)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (170)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (175)
OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-8742

Alignment Scores:

Pred. No.:	4.28e-12	Length:	188
Score:	168.00	Matches:	45
Percent Similarity:	73.77%	Conservative:	0
Best Local Similarity:	73.77%	Mismatches:	13
Query Match:	14.84%	Indels:	3
DB:	10	Gaps:	0

US-09-904-568-4 (1-224) x US-09-783-590-8742 (1-188)

QY 34 AlaAlaMetAlaArgLeuLeuGlyAspLeuAspArgSerThrPheArgLysLeuLeuLys 53
|||
Db 12 GCAGCAATGCCCGGCTACTAGGGACCTNAGACATNANACCGTTNAGAAAGTTGCTGAAG 71
QY 54 PheValValSerLeuGlnGlyGlu-AspCysArgAspGlyValGlnArgLeuGlyVa 73
|||
Db 72 TTTGTGTCAGCAGCTGCGAGGGGAGGAGCTGCCAAAGGTGTGCCAGGCTTTGGGT 131
QY 73 lSerAlaAsnLeuProGlu-GluGlnLeuGlyAlaLeuLeuAlaGly-MethHisThr 91
|||
Db 132 CAGCCCAACCTGNCGGAGNNGNCAGCTGGTGGCTGCTGNTGGANAGTCATGCACACA 188

RESULT 11

US-09-815-242-7826
Sequence 7826, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
LOCATION: (3)
OTHER INFORMATION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242

```

CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 7826
LENGTH: 945
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(945)
US-09-815-242-7826

Alignment Scores:
Pred. No.: 0.0138 Length: 945
Score: 97.00 Matches: 59
Percent Similarity: 39.61% Conservative: 23
Best Local Similarity: 28.50% Mismatches: 69
Query Match: 8.57% Indels: 56
DB: Gaps: 10

US-09-904-568-4 (1-224) x US-09-815-242-7826 (1-945)
QY 5 GYAAATATPrProTyrLeuHisHisProGluYAspSerHisSer----- 19
Db 229 GCGTTCGCGGATGCTGCTCGGGTCCGCGGCGAAACCCGACGACATCACTGCGCC 288
QY 20 -----GlyArgValSerPheLeuGluYAlaGlnLeuPro---ProGluValAlaAla 35
Db 289 GAGCGCGCGCGCGCTGACGACCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
QY 36 MetAlaAglLeuLeuGluYAspLeuAspArgSerThrPheArgValLeuLeuLeuYspheVal 55
Db 349 CAGCGCGCGCTGCTGAAAGCTGCGCGCTTGGCGCGGAGCAGACGCTGCGCGTGTGTC 408
QY 56 ValSerSerLeuGlnGluYAlaAspCysArgAspGlyValGlnArgLeuGlyValSerAla 75
Db 409 GCCGCGACCTGCGC-----CGCGGGTCTGAG-----CGCGGGTCTGAG 435
QY 76 AsnLeuProGluGlnGluYAlaLeuLeuAlaGlyMetHisThrLeu----- 92
Db 436 -----CGGAAATGCTGGGCGAACTGCTGCGCGCGCTTGGCGCGCTGCGCGCTGAAG 486
QY 93 -----LeuGlnGlnAlaLeuArgLeuProProThrSer 103
Db 487 GTCCGCTTCGACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
QY 104 LeuYspProAspThrPheArgAspGlnLeuGlnLeuYsp---LleProGlnAspLeu 122
Db 547 ATCAAGCCGGAAGCTC-----GAGGAATGCGCGCATCTCTGCTGCGCGCGCGCGATGAT 597
QY 123 ValGluAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSerValAla 142
Db 598 -----GACCTGTGTGCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
QY 143 Gln-----GlnGlnGlyAla-----TripleuProHisValAlaAspPhe 155
Db 652 CAGGTGTGATTTTCCCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 705
QY 156 ArgTrrArg-----ValAspValAlaIleSerThrSerAlaLeuAlaArg 170
```

```

Db 706 GCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 762
QY 171 SerLeuGlnProSerValLeu 177
Db 763 TCCCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 783

RESULT 12
US-09-848-915-2
Sequence 2, Application US/09848915
Patent No. US20020037518A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Goli, Surya K.
TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/848,915
FILING DATE: 04-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/822,260
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0247 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 751 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: UTRSN0702
CLONE: 2267574
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-848-915-2

Alignment Scores:
Pred. No.: 0.026 Length: 751
Score: 93.50 Matches: 32
Percent Similarity: 40.97% Conservative: 27
Best Local Similarity: 22.22% Mismatches: 60
Query Match: 8.26% Indels: 25
DB: Gaps: 4

US-09-904-568-4 (1-224) x US-09-848-915-2 (1-751)
QY 85 LeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuProProThrSerLeu 104
Db 232 CTAAGCGCAGGAAAGCCCGAGCTGACAGTCAACTTAAGCACTTATCTAAGAACTGT 291
QY 105 LysProAspThrPheArgAspGlnLeuGlnGluYspYsIleProGlnAspLeuValGly 124
Db 292 AAATTGAC-----ACAGAGCGCAATGAACTGTTTGGACGGAATATCAAGATTAATAG 345
```



```

QY 147 AlaTrpLeuProHisValAlaAspPheArg----- 156
Db 394 -----GTTGAAAGTTCGGCGAGAGAAATTCTGCTCCCTGTAAGCTAGAG 438
QY 157 -----TTPArgValAspValAlaIleSerThrSerAlaLeuAlaArgSerLeuGln 173
Db 439 ACYCTTGATGGCAGCTTAACTTACAGATGAGCTCCTCTCAAGCAAACTTAAATCT 498
QY 174 ProSerValLeuMetGlnLeuLysLeuSerAspGlySerAlaTyrArgPheGluValPro 193
Db 499 CCTCAAGCTCTGTATACAACTCGAGTGAACATGAAGATTCAAGAGCCCTGAGAAAGTT 558
QY 194 ThrAlaLysPheGlnGluLeuArgTyrSerValAlaLeuValLeuLysGluMetAlaAsp 213
Db 559 CTGTGTGAATTCAGTAC-----AAGGAGTTGTTGAT 591
QY 214 LeuGluLysArgCysGlu 219
Db 592 TTCTATACAAAGCTAGAG 609

RESULT 15
US-09-870-162A-22
/ Sequence 22, Application US/09870162A
/ Patent No. US2002004211B1
/ GENERAL INFORMATION:
/ APPLICANT: Breinig, Sabine
/ TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
/ FILE REFERENCE: BCI006 US Div
/ CURRENT APPLICATION NUMBER: US/09/870,162A
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: 09/516914
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 22
/ LENGTH: 1749
/ TYPE: DNA
/ ORGANISM: Thauera aromatica
US-09-870-162A-22

Alignment Scores:
Score: 0.265 Length: 1749
Seed: 89.50 Matches: 84
Percent Similarity: 31.79% Conservative: 26
Best Local Similarity: 24.28% Mismatches: 91
Query Match: 7.91% Indels: 145
Db: 10 Gaps: 19

US-09-904-568-4 (1-224) x US-09-870-162A-22 (1-1749)
QY 4 ValGlyAlaAlaThrProTyrLeuHisAspProGlyAspSerHisSerGlyArg----- 21
Db 580 GTTCGGCCCTGAAACGGCTTCTCTTCACTGAGGCGAGCTGGCATTTGCTGCGCGAC 639
QY 22 -----ValSerPheLeu 25
Db 640 GAGCGGACGTGATACGCGGGCTTGTTCATCCGCAAGCTGACCTGACCTTCTCTG 699
QY 26 GlyAlaGlnLeuProGluValAlaIleMetAlaArgLeuGluValAspLeuAspArg 45
Db 700 GAG-----CACCCGACGACCTTGTCTGCGCGCTCAG----- 732
QY 46 SerThrPheArgLysLeuLeuLysPheVal-----ValSerSerLeu 59
Db 733 -----CGGCAGATGCCGATTTCTCTCCAGAAAGGCTGCCGCGCTGCTC 783
QY 60 GlnGlyLysAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGlu 79
Db 784 -----GAACCTGCGCGGGCGGCTGCGAGCGCGGCTGTATGCCCTTTTCCAGCC 837
QY 80 GluGlnLeuGly----- 83

```

```

Db 838 GAAAGCTCGGCTCCACAGAGTCTGCGGAATCCGGCGCATGTTGATCCACACCGCG 897
QY 84 AlaLeuLeuAlaGlyMetHis----- 90
Db 898 GCGATCGCGCGCGCGCTCATTTCTGTACAGCGCGCGGTGAAGACTTCTTCCGACCCC 957
QY 91 -----ThrLeuGlnGlnAlaLeu 97
Db 956 CGTTTCGCCGACACCAAAATGGCATCTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTG 1017
QY 98 ArgLeuProPro-----ThrSerLeuLysProAspThrPheArg 110
Db 1018 CCGCAACCGGAACCGCGGAGTCGCGCATGATGCGCAGACAGCCGCGCGCTTATC 1077
QY 111 Asp-----GlnLeuGlnGluLysLysLeuProGlnAspLeuVal----- 123
Db 1078 GATCGTGAGCGCACTTCTTCCACTGCGCTGCGAGTGCATGATGCTGCGCGCATCGA 1137
QY 124 -----GlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeu----- 138
Db 1138 TTCGGGGCGGAACAGGCTCTGTATTCCTCGCATC-----CGCCATTCCTGCGCGCTT 1194
QY 139 ---AspSerValAla-----ValAlaAsp----- 154
Db 1195 GCCGACGATGGCAATGCTCGCACCCATGCCCGCATCTGCTTCTGTACAGAT 1254
QY 145 GlnGlyAlaTrpLeuProHis-----ValAlaAsp----- 154
Db 1255 CGGTG-----CCCATGAAGCGCGGTGAGTACCCGAGCATAGCCGATCTGGGT 1305
QY 155 -----PheArgTTPArgValAspValAlaIleSerThrSerAlaLeuAlaArg 170
Db 1306 CCAGCACACCGGTTCGAGCAGAGTTCCGAAGTGGCGTTGGAGACGCCCTCCACGA 1365
QY 171 SerLeuGlnProSerValLeu-----MetGlnLeu 180
Db 1366 ---ATTGATCCAGCGGAATCTGCGATTGAAGTCCGCGCGCGCGGTGCAATTCAGCTG 1422
QY 181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLeu 200
Db 1423 GAGCGGATGACCGCGCAATGCGCTGAGCGCTGACGTGCGCGCGCGGTGATGATTC 1482
QY 201 ArgTyrSerValAla-----LeuValLeuLysGluMetAlaAspLeuGluLysArg 217
Db 1483 CGCAAGTCTGCTGCGCGGTCTGTATCCGTGCTGCGCGAGCTCCGATCAGCGGACCGGA 1542
QY 218 CysGluArgArgLeuGln 223
Db 1543 TCGCAACCCCATGCGCAG 1560

```

Search completed: December 7, 2002, 04:24:24
 Job time : 39.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 7, 2002, 04:23:00 ; Search time 1551 Seconds
(without alignments)
2339.000 Million cell updates/sec

Title: US-09-904-568-4

Perfect score: 1132

Sequence: 1 MSAGVATPYLHPHGDHSG.....ALVLKEMADLEKRCERRLQD 224

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+pn2n.model -DEV=xlsp
-Q/cgn2.1/USPTO.spool/US0904568/runat_25112002_143305_19931/app_query.fasta_1.782
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0904568 @CGN 1 1 1716 @runat_25112002_143305_19931 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	93.6	905	13	BI837049
2	1052.5	93.0	1131	14	BM924244
3	1051	92.8	924	9	AL556701
4	1048	92.6	1115	12	BF795570
5	1010	89.2	846	12	BF792887
6	1003	88.6	716	13	BI752248
7	999	88.3	959	12	BG676661
8	995	87.9	870	13	BI835029
9	961	84.9	955	12	BG023805
10	961	84.9	975	14	BQ680339
11	953	84.2	670	14	BM849291
12	921	81.4	735	13	BI753202
13	921	81.4	936	12	BG290847
14	903	79.8	761	13	BI159794
15	893	78.9	999	11	AK010045
16	871	76.9	972	13	BI654666
17	869.5	76.8	797	13	BI870351
18	868	76.7	684	9	AU129385
19	867	76.6	1014	12	BG035515
20	858	75.8	867	10	BE277107
21	857	75.7	646	12	BG705627
22	857	75.7	1040	12	BF525462
23	852	75.3	829	12	BG422963
24	849.5	75.0	540	17	AQ055548
25	843.5	74.5	1093	13	BM423066
26	843	74.5	1081	13	BM545741
27	835.5	73.8	777	12	BF205159
28	829	73.2	600	13	BM314030
29	826.5	73.0	851	13	BI104334
30	817.5	72.2	521	14	BM849681
31	814	71.9	1711	11	AK013148
32	800	70.7	884	13	BI837085
33	792	70.0	727	14	BQ746289
34	787	69.5	726	14	BQ445783
35	781	69.0	1010	13	BI757152
36	772	68.2	502	14	BM694888
37	771.5	68.2	956	12	BE910440
38	768	67.8	891	12	BF685514
39	758	67.0	701	14	BM974637
40	758	67.0	1006	13	BM451357
41	756.5	66.8	676	10	BE390866
42	751	66.3	944	12	BG175412
43	749	66.2	704	9	AU130228
44	740	65.4	851	13	BM013651
45	726	64.1	610	12	BG754226

ALIGNMENTS

RESULT 1
BI837049
LOCUS
DEFINITION 603090003F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5228905 5',
905 bp mRNA linear EST 04-OCT-2001
rna sequence.
ACCESSION BI837049.1 GI:15948599
VERSION BI837049
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 905)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph. D.
Email: cgsabrs-remail.nih.gov
Tissue/Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLAM1576 row: e column: 02
High quality sequence stop: 888.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5228905"
/clone_1lb="NIH_MGC_120"
/1ab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally clone
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."

```

```

1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
Db 247 ATGCTGCTGGGAGCTGCAACTCAATACCTGCAATCATCTCTGGTGAATAGTCACAGTGGC 306

21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db 307 CGAGTGAGTTCTTCGGGGCCCGAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 366

41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db 367 GGGGACCTAGACAGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCTGCAG 426

61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db 427 GGGGAGGACTGCCAGAGGCTGTGAGGCTCTTGGGTGAGGCCCAACTGCCGAGGAG 486

81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 487 CAGCTGGGTGCCCTGTGGCAGGCATGCACACACTGCTCCAGCAGGCCCTCCGCTCTGCC 546

101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln 120
Db 547 CCCACAGCCTGAAGCCTGACACCTTCAGGACCAGCTCCAGGAGCTCTGCATCCCCCAA 606

121 AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer 140
Db 607 GACCTGGTGGGAGCTTGGCCAGCGTGTATTTGGGAGCCAGCGGCCCTCTCTGATCT 666

141 ValAlaGlnGlnGlnAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp 160
Db 667 GTGGCCACAGCAGCGGGCCCTGGCTGCCGCATGTTGCTGACTTTTCGGTGGCGGTGGAT 726

161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 727 GTAGCAATCTCCACCAAGTCCCTGGCTGCTCCCTCCAGCCAGCGCTCTGATGAGCTG 786

181 LysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLe 200
Db 787 AGCTTTTCAGATGGGTGTCAGCATACCCGCTTTGAGTTCCTCCACAGCAAGTTCCAGAGCT 846

200 uArgTyrSerValAlaLeuVal-LeuLysGluMetAlaAspLeu---GluLysArgCysG 219
Db 847 GCGGTACAGCTGGCCCTGGTCCCTTAAGAGATGCGAGATCTGGGANAAGGTGTGG 906

219 LuArgArgLeuGlnAsp 224
Db 907 ACCGCAAACTGCAGGAC 923

RESULT 3
AL556701 924 bp mRNA linear EST 16-FEB-2001
LOCUS AL556701 LTI_NFL006_PL2 Homo sapiens cdna clone CSODK005YG02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL556701
VERSION AL556701.1 GI:12899619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 924)
AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
CONTACT Genoscope
COMMENT Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..924
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="CSODK005YG02"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 163 a 265 c 316 g 179 t 1 others
ORIGIN

Alignment Scores: 7.1e-110 Length: 924
Score: 1051.00 Matches: 219
Percent Similarity: 97.78% Conservative: 1
Best Local Similarity: 97.33% Mismatches: 4
Query Match: 92.84% Indels: 3
DB: 9 Gaps: 0

US-09-904-568-4 (1-224) x AL556701 (1-924)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
Db 232 ATGCTGCTGGGAGCTGCAACTCAATACCTGCAATCATCTCTGGTGAATAGTCACAGTGGC 291

21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db 292 CGAGTGAGTTCTTCGGGGCCCGAGCTTCTCCAGAGGTGGCAGCAATGGCCGGCTACTA 351

41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db 352 GGGGACCTAGACAGAGCAGCAGTTCAGAAAGTTGCTGAAGTTTGTGGTCAGCAGCTTCAG 411

61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db 412 GGGGAGGACTGCCAGAGAGCTGTGAGCGCTCTTGGGTGAGGCCCAACTGCCGAGGAG 471

81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 472 CAGCTGGGTGGCTCTGTCGAGGAGCATGCACACACTGCTCCAGCAGGCCCTCCGCTCTGCC 531

101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLeuCysIleProGln 120
Db 532 CCCACAGCCTGAAGCCTGACACCTTCAGGACCAGCTCCAGGAGCTCTGCATCCCCCAA 591

121 AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer 140
Db 592 GACCTGGTGGGAGCTTGGCCAGCGTGTATTTGGGAGCCAGCGGCCCTCTCTGATCT 651

141 ValAlaGlnGlnGln-GlyAlaTrpLeuProHisValAlaAspPheArgTrpArgValAs 160
Db 652 GTGGCCACAGCAGCAGGGGGCCCTGCTGCGCATGTTGCTGACTTTTCGGTGGCGGTGGA 711

160 pValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLe 180
Db 712 TGTAGCAATCTCCACAGTGGCTTGGCTCGCTCCCTGAGCCAGCGAGCGTCTGATGAGCT 771

180 uLysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLe 200
Db 772 GAAGCTTTCAGATGGGTGAGCATACCGCTTTGAGGTCCCCCAGACCAAGTTCCAGAGCT 831

200 uArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArgCysGluAr 220
Db 832 GCGGTACAC-GTGGCCCTGGTCTCTAAAGAGATGGCAGATCTGAGAAAGARGTGTGAGCC 890

220 gArgLeuGlnAsp 224
Db 891 -AGACTGTCAGGAT 902

```

RESULT 4
LOCUS BF795570 1115 bp mRNA linear EST 12-JAN-2001
DEFINITION 602259634F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4342616 5',
mRNA sequence.
ACCESSION BF795570
VERSION BF795570.1 GI:12100716
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1115)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9958 row: d column: 09
High quality sequence stop: 708.
Location/Qualifiers
1..1115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4342616"
/clone_lib="NIH_MGC_85"
/issue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 259 a 277 c 377 g 202 t
ORIGIN
Alignment Scores:
Pred. No.: 2.16e-109 Length: 1115
Score: 1048.00 Matches: 212
Percent Similarity: 97.31% Conservative: 5
Best Local Similarity: 95.07% Mismatches: 4
Query Match: 92.58% Indels: 2
DB: 12 Gaps: 0
US-09-904-568-4 (1-224) x BF795570 (1-1115)
QY 1 MetSerAlaValAlGlyAlaAlaThrProTyrLeuHisHisProGlyAaspSerHisSerGly 20
Db 81 ATGCTGCTGCGGGGGCGCAACTCCATCACTGATCCTGCTGATGATGACACAGTGGC 140
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProProGluValAlaAlaMetAlaArgLeuLeu 40
Db 141 CGAGGAGTTCCTTGGGGGCCACGCTTCTCCAGAGGGGAGGACAAATGGCCCGGCTACTA 200
QY 41 GlyAspLeuAaPArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db 201 GGGGACCTAGACAGAGACACTTCAAGAACTTCTGGAAGTTTGGTACAGAGCTGACG 260
QY 61 GlyAluAaPArgSerGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGln 80
Db 261 GGGGAGAGCTGCCAGAGAGCGTGTGACGCTTGGGGCTACAGCCCAACTGCGGAGAG 320
QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 321 CAGCTGGGTGCTGCTGGCGAGGATGACACACTGCTCCAGAGGCCCTCGTCTGCC 380

QY 101 ProThrSerLeuLysPArgSerThrPheArgAaspGlnLeuGlnLeuCysAlaLeuProGln 120
Db 381 CCCACACAGCTGAAAGCTGACACCTTCAAGGACCAAGCTCCAGAGCTTGATGCCCA 440
QY 121 AspLeuValGlyAaspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAaspSer 140
Db 441 GACCTGTGGGGAGACTTGGCCAGAGCTGTATTTGGAGCCAGCGGCCCTCTGATTTCT 500
QY 141 ValAlaGlnGlnGlnGlyAlaTyrLeuProHisValAlaAaPArgArgTyrArgValAasp 160
Db 501 GTGGCCAGAGAGAGGGGGCTGGCTGCGCATGTCTTGCATCTTCCGTGGCGGTGAT 560
QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 561 GTAGCAATCTCCACAGAGCCCTGCTGCTCCCTGACAGCCAGAGCTGATGACAGCTG 620
QY 181 LysLeuSerAaspGlySerAla-TyrArgPheGluValProThrAlaLysPheGlnGlnLe 200
Db 621 AAGCTTTCAGATGGGTGACATTAACCGCTTGAAGTCCCAAGCAAGTTCCAGAGAGCT 680
QY 200 uArgTyrSerValAlaLeuValLeuLysGluMetAlaAaspLeu-GluLysArgCysGluA 220
Db 681 CGGTAAGAGCGTGGCCCTGCTTAAAGAGATGGGAATCTGGCAACAAGTGTGAAC 740
QY 220 TGAATG 221
Db 741 GCAGA 745
RESULT 5
LOCUS BF792887 846 bp mRNA linear EST 12-JAN-2001
DEFINITION 602253318F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4345628 5',
mRNA sequence.
ACCESSION BF792887
VERSION BF792887.1 GI:12097872
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 846)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9966 row: a column: 21
High quality sequence stop: 714.
Location/Qualifiers
1..846
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4345628"
/clone_lib="NIH_MGC_84"
/issue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 140 a 246 c 287 g 173 t
ORIGIN
Alignment Scores:

Pred. No.: 3,09e-105 Length: 846
 Score: 1010.00 Matches: 210
 Percent Similarity: 93.39% Conservative: 2
 Best Local Similarity: 92.51% Mismatches: 12
 Query Match: 89.22% Indels: 3
 DB: 12 Gaps: 0

US-09-904-568-4 (1-224) x BF792887 (1-846)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
 Db 95 ATGTCTGTCTGGGGGTCGCAACTCCATACCTGCATCATCTCTGGGTAGTACACAGTGGC 154
 Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaAlaArgLeuLeu 40
 Db 155 CGAGTGAGTTCTTGGGGGCCCGACCTTCTCCAGAGGTGGCAGCAATGGCCCGGTACTA 214
 Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
 Db 215 GGGGACCTAGACAGGACGACGTTTCAGAAAGTTGCTGAAGTTTGTGTCAGCAGCCTGCAG 274
 Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
 Db 275 GGGGAGGACTGCCGAGAGCTGTGAGCGTCTTGGGGTCAGCGCCCAACCTGCCGAGGAG 334
 Qy 81 GlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db 335 GAGCTGGGTGCCCTGCTGGCAGCATGCACACTGCTCCAGCAGGCCCTCCGTCGTGCC 394
 Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120
 Db 395 CCCACAGCCTGACAGCTTACACCTTCCAGGACAGCTCCAGGAGCTGTGCATCCCCAA 454
 Qy 121 AspLeuValGlyAspLeuAlaSerValVal-PheGlySerGlnArgProLeuLeuAspSe 140
 Db 455 GACCTGGTGGGGGACTTGGCCAGCGTGTATTTTGGGAGCCAGCGCCCTCTTGAATTC 514
 Qy 140 rValAlaGlnGlnGlnGlnValAlaTrpLeuProHisValAlaAspPheArgTrpArgValAs 160
 Db 515 TGTGCCCCAGCAGAGGGGGCTGGCTGCCGATGTTGCTGACTTTCGGTGGCGGTGA 574
 Qy 160 pValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGln-L 180
 Db 575 TGTAGCAATCTCCACAGTGCCCTGGCTCGCTCCCTGCAGCGAGCGCTCTGATGACGC 634
 Qy 180 euLysLeuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluL 200
 Db 635 TGAAGCTTTCAGATGGGTGAGCATACCGCTTTTGGAGGTCCCAAGTTCAGGAGC 694
 Qy 200 euArgTyrSerValAlaLeuValLeuLysGluMetAlaAspLeuGluLysArg-CysGln 219
 Db 695 TCGGTACACGGTGGCTGGTGGTCCCTAAGGAGATGGCGATCTCGGAAGAGGGTGTGAG 754
 Qy 220 ArgArgLeuGlnAsp 224
 Db 755 CGCAACTTGGGTGAT 769

RESULT 6
 BI752248
 LOCUS 603022442F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5192794 5',
 DEFINITION mRNA sequence.
 ACCESSION BI752248
 VERSION BI752248.1 GI:15743826
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 716)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cchapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LLAM11482 row: d column: 11
 High quality sequence stop: 716.

FEATURES

source
 1..716
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5192794"
 /clone_lib="NIH_MGC_114"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
 male brains, age range 23-27 yo. Library is oligo-dr
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.5 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 019. Note:
 this is a NIH MGC Library."

BASE COUNT 124 a 219 c 234 g 139 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,48e-104 Length: 716
 Score: 1003.00 Matches: 198
 Percent Similarity: 99.50% Conservative: 1
 Best Local Similarity: 99.00% Mismatches: 1
 Query Match: 88.60% Indels: 0
 DB: 13 Gaps: 0

US-09-904-568-4 (1-224) x BI752248 (1-716)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
 Db 117 ATGTCTGTCTGGGGGTCGCAACTCCATACCTGCATCATCTCTGGGTAGTACACAGTGGC 176
 Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
 Db 177 CGAGTGAGTTCTTGGGGGCCCGACCTTCTCCAGAGGTGGCAGCAATGGCCCGGTACTA 236
 Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
 Db 237 GGGGACCTAGACAGGAGCAGCTTTCAGAAAGTTGCTGAAGTTTGTGTCAGCAGCCTGCAG 296
 Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
 Db 297 GGGGAGGACTGCCGAGAGCTGTGAGCGTCTTGGGGTCAGCGCCCAACCTGCCGAGGAG 356
 Qy 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db 357 GAGCTGGGTGCCCTGTCGAGGCGATGCACACTGCTCCAGCAGGCCCTCCGTCGTGCC 416
 Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120
 Db 417 CCCACAGCCTGAAGCCTGACACCTTCCAGGACAGCTCCAGGAGCTGTGCATCCCCAA 476
 Qy 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
 Db 477 GACCTGGTGGGGGACTTGGCCAGCGTGTATTTGGGAGCCAGCGGCCCTCTCTTGTATTC 536
 Qy 141 ValAlaGlnGlnGlnGlnAlaTrpLeuProHisValAlaAspPheArgTrpArgValAsp 160
 Db 537 GTGGCCACAGCAGGAGGGGCTGGCTGCCGCGCATGTTGCTGACTTTCGGTGGCGGGTGGAT 596
 Qy 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180

Db 597 GTAGCAATCTCCACAGGCGCTGCTGCTCCCTGCAGCCAGCGCTCTGATGAGCTG 656
Oy 181 LysleuSerAspGlySerAlaTyArpPhegIuValProThAlaAlaPhegInglu 200
Db 657 AAGCTTCAGATGGGTCAAGATACCGCTTTGAGGTCCCAAGTCCAGAGAGCTG 716

RESULT 7
Bg676661 959 bp mRNA linear EST 01-MAY-2001
LOCUS 602623165F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4748057.5',
DEFINITION mRNA sequence.
ACCESSION Bg676661
VERSION Bg676661.1 GI:13908058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 959)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LHAM10599 row: a column: 18
High quality sequence start: 7
High quality sequence stop: 761.
Location/Qualifiers
1..959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4748057"
/clone_1id="NCI CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: Nci1;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 203 a 295 c 283 g 178 t
ORIGIN

Alignment Scores:
Pred. No.: 6.99e-104 Length: 959
Score: 999.00 Matches: 202
Percent Similarity: 98.54% Conservative: 1
Best Local Similarity: 98.06% Mismatches: 2
Query Match: 88.25% Indels: 1
DB: 12 Gaps: 0

US-09-904-568-4 (1-224) x Bg676661 (1-959)

Oy 20 GIATGValSerPheLeuGIValaGInleuProProGIuValAlaAlaMerAlaArgleu 39
Db 1 GGGCCAGTAGATTCTTGGGGGCCAGGCTCTCCAGAGGTGGCAGCAATGCCCCGGCTA 60
Oy 40 LeuGIAspLeuAspArgSerThrPheArglyLeuLeuLyPheValValSerSerleu 59
Db 61 CTAGAGGAGCCTAGACAGAGACGCTTCAGAAAGTTGCTGAAGTTGTGTCAGAGCTG 120
Oy 60 GInGIyGIuAspCyArAspGIyValaGInaGLeuGIyValSerAlaAsnleuProGIu 79
Db 121 CAGGGGGAGGACTGCTGGAGGCTGTGCAAGCTCTTGGGGTCAGGCCCAACTGCCGGAG 180
Oy 80 GIuGIuLeuGIyAlaLeuLeuAlaGIyMeThsThrLeuLeuGIuGInAlaLeuArgleu 99

Db 181 GAGCAGCTGGGGTCCCTGCTGCGAGGCATGCACACTGCTCCAGCAGGCCCTCGCTG 240
Oy 100 ProProThSerLeuLyProAspThrPheArgAspGInleuGInleuCysAllePro 119
Db 241 CCCCCCAGCAGCTGAGACCTTCAGAGACAGCTCCAGAGACTTGCATCCCC 300
Oy 120 GInAspLeuValGIyAspLeuAlaSerValaPheGlySerGInArgProleuLeuAsp 139
Db 301 CAAGACTGTCGCGGACTTGCCAGCGCTGTATTGGAGCCAGCGGCCCTCTGAT 360
Oy 140 SerValaGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 159
Db 361 TGTGTGGCCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Oy 160 AspValaAlaSerThrSerAlaLeuAlaArgSerLeuGInProSerValleuMeGIn 179
Db 421 GATGAGCAATCTCCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Oy 180 LeuLyLeuSerAspGlySerAlaTyArpPhegIuValProThAlaAlaPhegInglu 199
Db 481 CTGAAGCTTTCAGATGGGTCAAGATACCGCTTTGAGGTCCCAAGTCCAGAG 540
Oy 200 LeuArgTySerValaAlaLeuValleuLeuLeuAlaAspLeuGInLyS-ArgCyGcl 219
Db 541 CTGCGGTACACGCGGCTGCTGCTTAAAGAGATGCGATCTGAGAGACGAGGTGGA 600
Oy 219 uArgArgleuGInAsp 224
Db 601 GCGCAGACTGCAGAGC 616

RESULT 8
B1835029 870 bp mRNA linear EST 04-OCT-2001
LOCUS 603089252F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5228197.5',
DEFINITION mRNA sequence.
ACCESSION B1835029
VERSION B1835029.1 GI:15946579
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 870)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LHAM1574 row: g column: 14
High quality sequence stop: 772.
Location/Qualifiers
1..870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5228197"
/clone_1id="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site_1: Nci1; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is Oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics

tracking code 025. Note: this is a NIH_MGC Library."

BASE COUNT 169 a 266 c 272 g 163 t

ORIGIN

Alignment Scores:

Pred. No.: 1.7e-103 Length: 870

Score: 995.00 Matches: 210

Percent Similarity: 93.89% Conservativity: 5

Best Local Similarity: 91.70% Mismatches: 9

Query Match: 87.90% Indels: 5

DB: 13 Gaps: 0

US-09-904-568-4 (1-224) x BI835029 (1-870)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyVasSerHisSerGly 20

Db 169 ATGTCTGTGTGGGGCTGCAACTCATACCTGCATCTCTGGGTAGTACACAGTGC 228

Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40

Db 229 CGAGTGAGTTCTTGGGGCCCGACCTTCTCCAGAGTGGCAGCAATGGCCGGTACTA 288

Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60

Db 289 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGTGTCAGCAGCCTGCAG 348

Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80

Db 349 GGGGAGGACTGCGGAGAGCTGTGACGCTCTTGGGGTTCAGCGCAACCTGCGGAGGAG 408

Qy 81 GlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100

Db 409 CAGCTGGTGCCCTCTGCGGCGATGCACACTGCTCCAGAGCCCTCGCTGAC 468

Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGlnLeuCysLeProGln 120

Db 469 CCCACGAGCTGAAGCCTGACACCTTCAGGAGCAGCTCCAGAGCTTCGATCCGCCAA 528

Qy 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140

Db 529 GACCTGGTGGGGACTTGGCCAGCGGTGATTGGGAGCCAGCGCCCTCTCTGATTCT 588

Qy 141 ValAlaGlnGlnGlnGlyAlaThrLeuProHisValAlaAspPheArgTyrArgValAs 160

Db 589 GTGGCCACAGCAGGGGCTGGCTTGGCGCATGTGCTGACTTTCGTGGGGGTGA 648

Qy 160 pValAlaLeuSerThrSerAlaLeuAla-ArgSerLeuGlnProSerValLeuMetGlnL 180

Db 649 TGTAGCAATCTCCACAGTGCCTGGCTCGATCCCTGAGCGGAGCGTCTGATGCAGC 708

Qy 180 euLysLeuSerAspGlySer-AlaTyrArgPheGluValProThrAlalysPheGlnGlu 199

Db 709 TGAAGCTTTCAGATGGGTCAAGCATACCGCTTGAGGTCCCGCCAGACCAAGTTCAGAG 768

Qy 200 LeuArg-TyrSer-ValAlaLeuValLeuLysGlnMetAlaAspLeuGluLysArgCysG 219

Db 769 CTGCGGTACAGCGATGGCCCTGTCTTAAGGAGATGGCAATCTGGAGAGAGGTGTG 828

Qy 219 luArgArgLeuGlnAsp 224

Db 829 AGCGCAATGATGGAA 845

RESULT 9

BG023805

LOCUS 60230318F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4394812 5',

DEFINITION mRNA sequence.

ACCESSION BG023805

VERSION BG023805.1 GI:12408770

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (Bases 1 to 955)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM10091 row: c column: 05

High quality sequence stop: 709.

Location/Qualifiers

1. 955

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4394812"

/clone_lib="NIH_MGC_88"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site 1: NotI; Site 2: SalI; Cloned unidirectionally;

oligo-dT primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 189 a 260 c 307 g 199 t

ORIGIN

Alignment Scores:

Pred. No.: 1.59e-99 Length: 955

Score: 961.00 Matches: 214

Percent Similarity: 95.15% Conservativity: 2

Best Local Similarity: 94.27% Mismatches: 7

Query Match: 84.89% Indels: 7

DB: 12 Gaps: 0

US-09-904-568-4 (1-224) x BG023805 (1-955)

Qy 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyVasSerHisSerGly 20

Db 87 ATGTCTGTGTGGGGCTGCAACTCATACCTGCATCTCTGGGTAGTACACAGTGC 146

Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40

Db 147 CGAGTGAGTTCTTGGGGCCCGACCTTCTCCAGAGTGGCAGCAATGGCCGGTACTA 206

Qy 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60

Db 207 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTGCTGAAGTTTGTGTGTCAGCAGCCTGCAG 266

Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80

Db 267 GGGGAGGACTGCGGAGAGCTGTGAGCGTCTTGGGTGAGCGCCCAACCTGCGGAGGAG 326

Qy 81 GlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100

Db 327 CAGCTGGGTGCCCTGTGCGGCGATGCACACTGCTCCAGAGCCCTCTCGTCTGCC 386

Qy 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGlnLeuCysLeProGln 120

Db 387 CCCACGAGCCTGAAGCCTGACACCTTCAGGAGCAGCTCCAGAGCTTCGATCCGCCAA 446

Qy 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140

Db 447 GACCTGGTGGGGACTTGGCGAGCTGTGAGGTATTGGGAGCGAGCGGCCCTC-CTTGATTCT 505

Qy 141 ValAlaGlnGlnGlnGlyAlaThrLeuProHisValAlaAspPheArgTyrArgValAsp 160

Db 506 GTGGCCACAGCAGGGGCTGGCTGCCGATGTTGCTGACTTTTCGGTGGCGGTGGAT 565

QY 161 ValAlaIleSerThrSerAla-LeuAlaArgSerLeuGlnProSerValIleuMetGlnLe 180
 Db 566 GTAGCAATCTCCACCAAGTCCCTCGCTCGCTCCGCGAGCGAGCGT CTGATCAGACT 624

QY 180 UlyleuSerAspGlySerAlaTyrArgPheGluValProThrAlaLysPheGlnGluLe 200
 Db 625 GAAGCTTTCAGATGGGTGAGCATACCGCTTTGAGGTCCCA-GGCAAGTTTCAAGAGCT 683

QY 200 UArgTyrSerValAlaLeuValIleuLysGlnMet-AlaAspLeuGlu-LysArgCysGlu 219
 Db 664 GCGGACAGCGCTGCGCTCGCTTAAGGAGATGGGACATTGGAGAGAGAGGTGTGAG 743

QY 220 -ArgArgLeuGln 223
 Db 744 CCGCCGACTGAGA 756

RESULT 10
 B0680339 975 bp mRNA linear EST 15-JUL-2002
 LOCUS AGENCOURT 8061618 NIH_MGC_112 Homo sapiens CDNA clone IMAGE:6214125
 DEFINITION 5', mRNA sequence.
 ACCESSION B0680339
 VERSION B0680339.1 GI:21793018
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 975)
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: DCM/DMP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LCM2380 row: C column: 22
 High quality sequence stop: 638.
 Location/Qualifiers
 1..975
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="6214125"
 /clone_lib="NIH_MGC_112"
 /issue_type="melanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAG(G). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH MGC Library."

BASE COUNT 187 a 301 c 319 g 168 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.65e-99 Length: 975
 Score: 961.00 Matches: 195
 Percent Similarity: 98.00% Conservative: 1
 Best Local Similarity: 97.50% Mismatches: 2
 Query Match: 84.89% Indels: 2
 DB: 14 Gaps: 0

US-09-904-568-4 (1-224) x B0680339 (1-975)

QY 1 MetSerAlaValGlyAlaAlaThrProTyrIleuHisHisProGlyAspSerHisSerGly 20
 Db 129 ATGTCTGCTGGGGGGGTGCAATCTCATCTGATCTCTGGGATAGTACAGTGGC 188

QY 21 ArgValSerPheLeuGlnValGlnLeuProProGluValAlaAlaMetAlaArgLeuLeu 40
 Db 189 CGAGTGAATTTCTTGGGGGGCCAGCTTCTCCAGAGGGGCGAGCAATGGCCCGGTACTTA 248

QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
 Db 249 GGGGACCTAGACAGAGAGACCTTCAAGAAAGTTGCTGAAGTTTGCTAGACAGCTTCGAG 308

QY 61 GlyIleuAspCysArgAspGlyValGlnArgLeuGlnValSerAlaAsnLeuProGlu 80
 Db 309 GGGGAGACTCCCGAGAGAGCTGTGCAGCGCTTGGGGGTCAAGCCCAACTCCGAGAGAG 368

QY 81 GlnLeuGlnAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
 Db 369 CAGCTGGGTGCTCTGTGGAGGAGATGCACACACTGCTCCAGAGGCGCTTCGTGGCC 428

QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnGluLeuCys11eProGln 120
 Db 429 CCCACACCTCGAAGCCTTGACACTTCAGGSAACAGCTCCAGAGCTTGATCCGCCA 488

QY 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
 Db 489 GACCTGTGCGGGGACTTGCGCAGCGTGGTATTTGGAGACCAGCGGCCCTCTTGATCT 548

QY 141 ValAlaGlnGlnGlnGlnValAlaThrLeuProHisValAlaAspPheArgTyrArgValAsp 160
 Db 549 GTGGCCAGCAGCAGAGGGGGCTGCTGCGCATGTGTGACTTTCGGTGGCGGGTGGAT 608

QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValIleuMetGlnLeu 180
 Db 609 GTAGCAATCTCCACCAAGTCCCTCGCTCGCTCCGCGAGCGAGCGTCTGATCAGCTG 668

QY 181 LysLeuSerAspGlySerAlaTyr-ArgPheGlu-ValProThrAlaLysPheGln 198
 Db 669 AAGCTTTCACATGGGTGATCAGCATACCGCTTTGAGAGGTCCACAGCCCAAGTTCCAG 724

RESULT 11
 B0849291 670 bp mRNA linear EST 06-MAR-2002
 LOCUS K-EST0129329 S13KMS5 Homo sapiens CDNA clone S13KMS5-67-G03 5',
 DEFINITION mRNA sequence.
 ACCESSION B0849291
 VERSION B0849291.1 GI:19205690
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 670)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongjung@mail.krdb.re.kr
 Plate: 67 row: G column: 03
 High quality sequence stop: 670.
 Location/Qualifiers
 1..670
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S13KMS5-67-G03"
 /clone_lib="S13KMS5"

FEATURES
 source


```
/tissue type="myeloma"
/cell_line="KMS-5"
/lab_host="top10p"
/note="Vector: pCNS; Site_1: EcoRI; Site_2: NotI; The poly
(A)+ RNA was dephosphorylated with bacterial alkaline
phosphatase (BAP) and then decapped with tobacco acid
pyrophosphatase (TAP). The decapped intact mRNA was
ligated with DNA-RNA linker including EcoR I site by
treatment of T4 RNA ligase and the first strand cDNA was
synthesized from oligo dT-selected mRNA by priming with
dT-tailed vector. The dT-tailed vector was adjusted to
have about 60nt. The cDNA vector was circularized with E.
coli DNA ligase after digestion of EcoRI which site is
also included in vector. An RNA strand converted to a DNA
strand by Okazama-Berg method. The obtained cDNA vectors
were used for transformation of competent cells E. coli
top10p' by electroporation method. The cDNA libraries
constructed by this method are full-length enriched cDNA
library."
BASE COUNT 117 a 203 c 217 g 133 t
ORIGIN
Alignment Scores:
Pred. No.: 7.19e-99 Length: 670
Score: 953.00 Matches: 188
Percent Similarity: 99.47% Conservative: 1
Best Local Similarity: 98.95% Mismatches: 1
Query Match: 84.19% Indels: 0
DB: 14 Gaps: 0
US-09-904-568-4 (1-224) x BM849291 (1-670)
QY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
DB 99 ATGTCGTGCTGGGGGCTGCACTCCATACCTGCATCATCTCTGGTATAGTCACATGGC 158
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
DB 159 CGAGTCAGTTCTTGGGGCCAGCTTCTCCAGAGTGGCAGCATGGCCGGCTACTA 218
QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
DB 219 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTTCTGAAGTTTCTGGTTCAGCAGCTGCAG 278
QY 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
DB 279 GGGGAGGACTGCCGAGAGGCTGTGCAGCGCTTTGGGGTTCAGCGCCAACTGCCGAGGAG 338
QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
DB 339 CAGCTGGTGCCCTGTGGGAGGATGCACACTGCTCCAGCAGGCCCTCCGTCTGCCC 398
QY 101 ProThrSerLeuLysProAspThrPheArgAspGlnLeuGlnLeuGlyCysIleProGln 120
DB 399 CCCACAGCCTGAAGCCTGACACTTCAGGAGCAGCTCCAGGAGCTCTGCATCCGCCAA 458
QY 121 AspLeuValGlyAspLeuAlaSerValValPheGlySerGlnArgProLeuLeuAspSer 140
DB 459 GACCTGGTCGGGACTTGGCGAGCGTGGTATTTGGGAGCGAGCGGCCCTCCCTGTATCT 518
QY 141 ValAlaGlnGlnGlnGlyAlaTTPLeuProHisValAlaAspPheArgTrpArgValAsp 160
DB 519 GTGGCCAGCAGCAGGGGGCTGGCTGCCCATGTTGCTGACTTTCGGTGGCGGGTGGAT 578
QY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
DB 579 GTAGCAATCTCCACAGTGCCTGGCTGCTCCCTGAGCGCGAGCGCTCTGTATGATCG 638
QY 181 LysLeuSerAspGlySerAlaTyrArgPhe 190
DB 639 AAGCTTTTCAGATGGGTACGATACCGCTTT 668
RESULT 12
```

```
BI753202
LOCUS BI753202 735 bp mRNA linear EST 25-SEP-2001
DEFINITION 603026082F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5196260 5',
mRNA sequence.
ACCESSION BI753202
VERSION BI753202.1 GI:15744780
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11491 row: d column: 21
High quality sequence stop: 731.
FEATURES
Location/Qualifiers
source
1..735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5196260"
/clone_lib="NIH MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 019. Note:
this is a NIH MGC Library."
BASE COUNT 128 a 239 c 233 g 135 t
ORIGIN
Alignment Scores:
Pred. No.: 3.95e-95 Length: 735
Score: 921.00 Matches: 182
Percent Similarity: 99.46% Conservative: 1
Best Local Similarity: 98.91% Mismatches: 1
Query Match: 81.36% Indels: 0
DB: 13 Gaps: 0
US-09-904-568-4 (1-224) x BI753202 (1-735)
QY 1 MetSerAlaValGlyAlaAlaThrProTyrLeuHisHisProGlyAspSerHisSerGly 20
DB 182 ATGTCGTGCTGGGGGCTGCACTCCATACCTGCATCATCTCTGGTATAGTCACATGGC 241
QY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
DB 242 CGAGTCAGTTCTTCTGGGGGCCAGCTTCTCCAGAGGTTGGCAGCAATGGCCCGCTACTA 301
QY 41 GlyAspLeuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
DB 302 GGGGACCTAGACAGGAGCAGCTTCAGAAAGTTTCTGAAGTTTCTGGTTCAGCAGCTGCAG 361
QY 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
DB 362 GGGGAGGACTGCCGAGGAGGCTGTGCAGCGCTTGGGGTTCAGCGCCAACTGCCGAGGAG 421
QY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
```

Db 422 CAGCTGGGTCCTGCTGGCAGGATGACACACTGCTCCAGAGGCGCTCCGTCTGCC 481
OY 101 ProthrsrLeuLysProAspThrPharArgAspGlnLeuGlnGluLeuCyS11eProGln 120
Db 482 CCCACAGCCTGAGAGCTGACACCTTCAGAGGACAGAGCTCCAGAGCTTCGATCCGCCA 541
OY 121 AspleuValGlyAspleuLysSerValValPheGlySerGlnArgProLeuLysAspSer 140
Db 542 GACCTGGTGGGACTGTGGCCAGCGGTGTATTTGGAGCCAGCGGCCCTCTTATTTCT 601
OY 141 ValAlaGlnGlnGlnGlnValATrPleuProH1sValAlaAspPheArgTrpArgValAsp 160
Db 602 GTGGCCCGACAGCAGAGGCGGCTGTGGCTCCGACATTTGCTGATTTTGGTGGGTGAT 661
OY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 662 GTAGCAATCTCCACAGAGTCCCTGCTCGCTCCCTGACAGCGAGCTGTATGACAGCTG 721
OY 181 LysLeuSerAsp 184
Db 722 AAGCTTCAGAT 733

RESULT 13
Bg290847 936 bp mRNA linear EST 21-FEB-2001
LOCUS Bg290847
DEFINITION 602389190f1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4517734 5',
mRNA sequence.
ACCESSION Bg290847
VERSION Bg290847.1 GI:13048222
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 936)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM10411 row: d column: 23
High quality sequence stop: 704.

FEATURES
Location/Qualifiers
1..936
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4517734"
/clone_1db="NIH MGC 93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-SPORT6, Site_1: Nci1;
Site_2: Sal1; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 186 a 263 c 321 g 166 t
ORIGIN

Alignment Scores:
Pred. No.: 5.96e-95 Length: 936
Score: 921.00 Matches: 202
Percent Similarity: 95.37% Conservative: 4
Best Local Similarity: 93.52% Mismatches: 5
Query Match: 81.36% Indels: 9
DB: 12 Gaps: 1

US-09-904-568-4 (1-224) x Bg290847 (1-936)

OY 1 MetSerAlaValAlaAlaThrProTyLeuH1sH1sProGlyAspSerH1sSerGly 20
Db 87 ATGTCTGCTGTGGGGCTGCAACTTCATCTGATCTCTGGTATGATACAGTGAC 146
OY 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40
Db 147 CGAGTAGATTTCTTGGGGCCAGGCTTCTCAGAGGTGGCAGCAATGGCCGGTACTTA 206
OY 41 GlyAspleuAspArgSerThrPheArgLysLeuLeuLysPheValValSerSerLeuGln 60
Db 207 GGGAGCCTAGACAGAGACACCTTCAAGAAAGTTGCTGAAGTTGTGTACAGACGCTGACG 266
OY 61 GlyLysAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80
Db 267 GGGAGGAGCTCCAGAGAGCTGTGACAGGCTTTGGGGTACAGCCACCTTCCGAGGAG 326
OY 81 GlnLeuGlyAlaLeuLeuAlaGlyMetH1sThrLeuLeuGlnGlnAlaLeuArgLeuPro 100
Db 327 CAGCTGGTGGCTGCTGTGGCAGCATGACACACTGCTCCAGAGGCGCTCCGTCTGCC 386
OY 101 ProthrsrLeuLysProAspThrPharArgAspGlnLeuGlnGluLeuCyS11eProGln 120
Db 387 CCCACAGCCTGAGAGCTGACACCTTCAGAGCAGCTCCAGAGCTTCGATCCGCCA 446
OY 121 AspleuValGlyAspleuLysSerValValPheGlySerGlnArgProLeuLysAspSer 140
Db 447 GACCTGGTGGGAGACTTGGCCAGCGTGATTT-GGAGCCAGCGGCGCTC-CTTGATTTCT 504
OY 141 ValAlaGlnGlnGlnGlnValATrPleuProH1sValAlaAspPheArgTrpArgValAsp 160
Db 505 GTGGCCCGACAGCAGAGGCGGCTGTGGCTGGCATTTGCTGATTTGCGTGGCGGTGAT 564
OY 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180
Db 565 GTAGCAATCTCCACAGAGTCCCTGCTCGCTCCCTGACGCGAGCGT-CTGATCAGCTG 623
OY 181 LysLeuSerAspGlySerAlaTyArgPheGlyValProThrAlaLysPheGlnGluLeu 200
Db 624 AAGCTTCAGATGGGTGACGATACCGCTT-GAGGTCCCGACAGCGAGCTTCAGAGCTG 682
OY 201 ArgTySerValAlaLeuValLeuGlyMetAlaAspleuGluLys 216
Db 683 CGGTACAGCGTGGCTGG--TTTAAAGAGATGACATTCGACGAA 727

RESULT 14
B1159794 761 bp mRNA linear EST 05-JUL-2001
LOCUS B1159794
DEFINITION 602863564f1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5017958 5',
mRNA sequence.
ACCESSION B1159794
VERSION B1159794
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 761)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1830 row: o column: 15
High quality sequence stop: 755.

FEATURES		Location/Qualifiers		DEFINITION	
source		1..761		Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:231006SH03:homolog to CDNA FLJ13008 FIS, CLONE NT2R3000456 (HYPOTHETICAL 24.7 KDA PROTEIN) (HT002 PROTEIN, HYPERENSION-RELATED CALCIUM-REGULATED GENE), full insert sequence. AK010045	
		/organism="Homo sapiens"		AK010045.1 GI:12845215	
		/db_xref="taxon:9606"		HTC; CAP trapper.	
		/clone="IMAGE:5017958"		Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone lib:RIKEN full-length enriched mouse cDNA library	
		/clone lib="NIH MGC 42"		clone:231006SH03.	
		/lab host="DH10B (phage-resistant)"		Mus musculus	
		/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. "		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
BASE COUNT		130 a 220 c 261 g 150 t		1	
ORIGIN				2	
Alignment Scores:				3	
Pred. No.:		4..878-93		Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
Score:		903.00		RIKEN integrated sequence analysis (RISA) system--384-format	
Percent Similarity:		97.87%		Genome Res. 10 (11), 1757-1771 (2000)	
Best Local Similarity:		97.34%		20530913	
Query Match:		79.77%		11076861	
Indels:		1		4	
Gaps:		0		Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Arakawa, T., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Rong, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S. and Hayashizaki, Y.	
US-09-904-568-4 (1-224) x B1159794 (1-761)				TITLE	
Qy 1 MetSerAlaValGlyAlaAlaThrProTyLeuHisHisProGlyAspSerHisSerGly 20				JOURNAL	
Db 199 ATGTCGTCTGTGGGACTGCACTTCACTTCATCTGCTATAGTACATGGC 258				MEDLINE	
Qy 21 ArgValSerPheLeuGlyAlaGlnLeuProGluValAlaAlaMetAlaArgLeuLeu 40				PUBMED	
Db 259 CGAGTCAGTTCCTTGGGGGCCAGCTTCCTCCAGAGTGGCAGCAATGGCCCGCTACTA 318				REFERENCE	
Qy 41 GlyAspLeuAspArgSerThrPheArgGlyLeuLeuLeuPheValValSerLeuGln 60				AUTHORS	
Db 319 GGGGACCTAGACAGGACGCTTCAGAAAGTTGCTGAGAGTTTGTGTGACGACCTGCAG 378				TITLE	
Qy 61 GlyGluAspCysArgAspGlyValGlnArgLeuGlyValSerAlaAsnLeuProGluGlu 80				JOURNAL	
Db 379 GGGGAGACTGCGGAGGCTGTGCGAGCTCTTGGGGTCCAGCCCACTGCGGAGGAG 438				MEDLINE	
Qy 81 GlnLeuGlyAlaLeuAlaGlyMetHisThrLeuLeuGlnGlnAlaLeuArgLeuPro 100				PUBMED	
Db 439 CAGCTGGTGGCTGCTGGCAGGATGCACACACTGCTCCAGCAGGCGCTCCGCTTGGC 498				REFERENCE	
Qy 101 ProThrSerLeuLeuProAspThrPheArgAspGlnLeuGlnGluLeuCysIleProGln 120				AUTHORS	
Db 499 CCACACGACCTGAAGCTTGACCTTCAGGAGACGCTCCAGAGAGCTCTGCATCCCCCAA 558				TITLE	
Qy 121 AspLeuValGlyAspLeuAlaSerValPheGlySerGlnArgProLeuLeuAspSer 140				JOURNAL	
Db 559 GACTGTGCGGAGCTGGCAGCGTGTATTTGGGAGCCAGCGGCCCTCTGATTCT 618				MEDLINE	
Qy 141 ValAlaGlnGlnGlnAlaTrpLeuProHisValAlaAlaAspPheArgTrpArgValAsp 160				PUBMED	
Db 619 GTGCGCCAGCAGCAGGGGGCTGGCTCCGATGTTGCTGACCTTTCGGTGGCGGTGGAT 678				REFERENCE	
Qy 161 ValAlaIleSerThrSerAlaLeuAlaArgSerLeuGlnProSerValLeuMetGlnLeu 180				AUTHORS	
Db 679 GTAGCAATCTCCACAGT-GCCCTGGCTGCTCCCTGTCAGCGGAGCGCTGTGATGACGTG 737				TITLE	
Qy 181 LysLeuSerAspGlySerAlaTy 188				JOURNAL	
Db 738 AAGCTTTCAGATGGTCCAGCATAC 761				MEDLINE	
RESULT 15				PUBMED	
AK010045				REFERENCE	
LOCUS				AUTHORS	

